A Primer on Quantum Channel Capacity and Its Applications to Optical Communication

Matthew Thill*, Sam Dolinar*, and Dariush Divsalar*

ABSTRACT. — In this article, we review the capacity of a quantum channel for transmitting classical and quantum information, with our primary focus being the context of optical communication. We discuss the necessary trade-offs between the rates of sending both bits and qubits over a quantum channel, and how these rates are affected in the presence of shared entanglement between sender and receiver. We also review several protocols which achieve these rates. Then, shifting our focus to the free-space optical channel, we review common quantum states of photons and methods for modulating information on them. We discuss the capacities associated with these modulation techniques as well as the prospect of feasibly generating and exploiting entanglement to boost the classical capacity of optical free-space communication.

I. Introduction

Due to the bosonic nature of photons, the free-space optical channel is an instance of a quantum channel. Not only can it be used to transmit classical information (measured in bits), but it can transmit quantum information (measured in qubits) and be used to establish shared entanglement (measured in ebits) between sender and receiver. Thus, in order to optimally utilize the free-space optical channel, it is necessary to study the trade-off between these three resources.

The various capacities of the general quantum channel have been studied for decades, though have yet to be completely characterized. Whereas a classical channel has a single associated capacity, a quantum channel has four: the classical capacity for transmitting bits, the quantum capacity for transmitting qubits, the entanglement-assisted classical capacity for transmitting bits while consuming ebits, and the private classical capacity

*Communications Architectures and Research Section

The research described in this publication was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.
© 2017 California Institute of Technology. Government sponsorship acknowledged.

for transmitting bits which remain secret from an eavesdropper.

In this article, we review what is known about the various capacities of the general quantum channel as well as the trade-offs between the rates at which a quantum channel can transmit classical and quantum information while consuming (or producing) shared entanglement between the sender and the receiver. We will discuss several known protocols for trading between these resources. We will then shift our focus to the lossy bosonic channel, which is the model for free-space optical communication.

It should be emphasized that this article is intended primarily for those with some background in information theory who require a primer on quantum optical communications. As such, Section II reviews some of the basic elements of quantum theory, including a discussion on quantum states, purifications, and measurements. Sections III and IV generalize familiar concepts and quantities from classical information theory to the quantum regime. Section III focuses on entropy and data compression, while Section IV discusses quantum mutual information and several related quantities, some of which do not have classical counterparts. Sections IV-C and IV-D also review the formal definition of a quantum channel and the notion of the various resources involved in quantum communication. Section V realizes several notorious protocols as trade-offs between these resources, including super-dense coding and teleportation.

Readers with more quantum background may wish to skip directly to Section VI, in which we formally define the four capacities of a quantum channel which we mentioned above, and summarize what is known about how to quantify or bound them. In Section VII, we discuss several achievable protocols which generalize those from previous sections. We broach on the use of "noisy" versions of the basic quantum resources, such as partially entangled states or noisy channels which do not maintain the coherence of qubits, and we briefly discuss how arbitrary entangled states can be converted to others, thus justifying the ability to quantify the amount of entanglement between states as a resource for quantum communication.

Finally, in Section VIII, we focus on optical communication, beginning with a review of common optical quantum states in Section VIII-A and the lossy bosonic channel—our model for the free-space optical quantum channel—in Section VIII-B. We also review the aforementioned capacities in this context, with an emphasis on entanglement-assisted classical communication, and we discuss the feasibility of boosting the rate of classical communication in a setting in which prior shared entanglement is limited. One important caveat to the earlier results about the capacities of the general quantum channel is that we typically assume access to any theoretically-permissible quantum states and receivers for use in communication. In practice, we are limited to what we can produce in the lab. Thus, in Section VIII-C, we review how well we can approach the classical capacity of the bosonic channel when restricting ourselves to several common types of photon states for communication.

Much of the material in Sections III-VII is substantially elaborated in Mark Wilde's

textbook *Quantum Shannon Theory* [1]. This is an invaluable resource, and we borrow heavily from the notation and terminology he establishes therein. Our hope is that these sections provide a means for the interested reader to quickly digest and gain a working knowledge of this material.

II. Quantum Basics

A. Quantum States and Composite States

We begin with several fundamental definitions. Let $\mathcal{H}$ be a complex Hilbert space representing a quantum state space. In much of our discussion, we will take $\mathcal{H}$ to be discrete and even finite dimensional with the understanding that generalizations are straightforward. We will consider two different types of quantum states. A *pure state* is a vector $|\psi\rangle \in \mathcal{H}$, where we have used the traditional bra-ket notation. As such, we denote the adjoint of $|\psi\rangle$ as $\langle\psi|$, and the inner product between two pure states $|\psi_1\rangle$ and $|\psi_2\rangle$ as $\langle\psi_1|\psi_2\rangle$. Likewise, their outer product may be written $|\psi_1\rangle\langle\psi_2|$. If we let $\{|i\rangle\}$ be an orthonormal basis for $\mathcal{H}$, where i indexes the basis elements, then assuming a pure state $|\psi\rangle$ is normalized, we interpret $|\langle i|\psi\rangle|^2$ as the probability that $|\psi\rangle$ evolves to state $|i\rangle$ when measured in the $\{|i\rangle\}$ basis. We will discuss the evolution and measurement of quantum states in greater detail shortly, but in light of this detail we will assume that our pure states are normalized, $|\langle\psi|\psi\rangle| = 1$, and we only distinguish pure states up to a phase offset (that is, we do not distinguish between $|\psi\rangle$ and $\alpha|\psi\rangle$ for $|\alpha| = 1$).

We may also consider an ensemble of pure states $\{p_i, |\psi_i\rangle\}$, $0 \leq p_i \leq 1$, $\sum_i p_i = 1$, which we call a *mixed state*, with corresponding density operator $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$. We can interpret ρ as a "noisy" quantum state. The density operators of this form are precisely the set of Hermitian nonnegative operators of trace 1 (see, for instance, [1]). Given a pure state $|\psi\rangle$, we will sometimes adopt the shorthand $\psi := |\psi\rangle\langle\psi|$ to represent its density operator (corresponding to the ensemble $\{1, |\psi\rangle\}$). Given a subspace $\mathcal{M} \subseteq \mathcal{H}$, the probability of observing a mixed state ρ in this subspace is $P(\rho, \mathcal{M}) := \mathrm{Tr}\left(\Pi_{\mathcal{M}}\rho\right)$, where $\Pi_{\mathcal{M}}$ is the projection operator onto $\mathcal{M}$.

Given two pure states $\psi_A \in \mathcal{H}_A$ and $\psi_B \in \mathcal{H}_B$, they can be thought of as sharing a quantum state ψ_{AB} in the tensor space $\mathcal{H}_A \otimes \mathcal{H}_B$. Similarly, two mixed states ρ_A and ρ_B share a state $\rho_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$. We call the ψ_{AB} and ρ_{AB} *composite states*, which live in the composite quantum system $\mathcal{H}_A \otimes \mathcal{H}_B$. For a composite state ρ_{AB}, the corresponding ρ_A and ρ_B are called *local operators*.

If a composite state ρ_{AB} can be decomposed in the form $\rho_{AB} = \sum_x p_x \rho_{A,x} \otimes \rho_{B,x}$ where $\{p_x\}$ forms a probability distribution, then we call the state *separable*. Otherwise, we call it *entangled*. Here, '$\otimes$' denotes the Kronecker product of operators. As shorthand, we will often denote the pure separable state $|\psi\rangle_A \otimes |\psi\rangle_B$ simply as $|\psi\rangle_A|\psi\rangle_B$. Separable states are statistically independent in the sense that, given subspaces $\mathcal{M}_A \subset \mathcal{H}_A$ and

$\mathcal{M}_B \subset \mathcal{H}_B$, we have

$$P(\rho_A \otimes \rho_B, \mathcal{M}_A \otimes \mathcal{M}_B) = \mathrm{Tr}\left(\Pi_{\mathcal{M}_A} \otimes \Pi_{\mathcal{M}_B}\ \rho_A \otimes \rho_B\right)$$
$$= \mathrm{Tr}\left(\Pi_{\mathcal{M}_A}\rho_A\right)\mathrm{Tr}\left(\Pi_{\mathcal{M}_B}\rho_B\right) = P(\rho_A, \mathcal{M}_A)P(\rho_B, \mathcal{M}_B). \quad (1)$$

That is, the probability of observing $\rho_A \otimes \rho_B$ in the space $\mathcal{M}_A \otimes \mathcal{M}_B$ is the product of the probabilities of observing the local operators in their respective subspaces.

In general, the composite operator ρ_{AB} cannot be determined from knowledge of the individual states ρ_A and ρ_B alone, but depends on the interrelation between them. The local states, however, can be recovered by taking the *partial trace* of the composite state: $\rho_A = \mathrm{Tr}_{\mathcal{H}_B}(\rho_{AB})$ and $\rho_B = \mathrm{Tr}_{\mathcal{H}_A}(\rho_{AB})$. The partial trace is defined as follows:

Definition 1 (Partial Trace)**.** *Let X_{AB} be a linear operator over $\mathcal{H}_A \otimes \mathcal{H}_B$. The partial trace of X_{AB} over $\mathcal{H}_B$ is the linear operator $Tr_{\mathcal{H}_B}(X_{AB})$ over $\mathcal{H}_A$ defined as*

$$Tr_{\mathcal{H}_B}(X_{AB}) := \sum_i (I_A \otimes \langle i|_B) X_{AB} (I_A \otimes |i\rangle_B), \quad (2)$$

where $\{|i\rangle_B\}$ is an orthonormal basis for $\mathcal{H}_B$. Equivalently, if X_{AB} can be written (in any way) as $\sum_m A_m \otimes B_m$, then $Tr_{\mathcal{H}_B}(X_{AB}) = \sum_m A_m\, Tr(B_m)$. The partial trace over $\mathcal{H}_A$, $Tr_{\mathcal{H}_A}(X_{AB})$ is defined similarly. We define the local operators $X_A := Tr_{\mathcal{H}_B}(X_{AB})$ and $X_B := Tr_{\mathcal{H}_A}(X_{AB})$, and use the phrase "tracing over $\mathcal{H}_B$" (resp. "$\mathcal{H}_A$") to refer to them.

Given a composite density operator ρ_{AB} with local operator $\rho_A = \mathrm{Tr}_{\mathcal{H}_B}(\rho_{AB})$, and a subspace $\mathcal{M}_A \subset \mathcal{H}_A$, we can immediately find the probability of observing ρ_A in space $\mathcal{M}_\mathcal{A}$ from the relation $P(\rho_A, \mathcal{M}_A) = \mathrm{Tr}\left(\Pi_{\mathcal{M}_A}\rho_A\right) = \mathrm{Tr}\left(\Pi_{\mathcal{M}_A \otimes \mathcal{H}_B}\rho_{AB}\right) = P(\rho_{AB}, \mathcal{M}_A \otimes \mathcal{H}_B)$.

B. Purifications and the Schmidt Decomposition

Given a composite pure state $|\psi\rangle_{AR} \in \mathcal{H}_A \otimes \mathcal{H}_R$ with corresponding operator $\psi_{AR} = |\psi\rangle\langle\psi|_{AR}$, the local operator $\rho_A = \mathrm{Tr}_{\mathcal{H}_R}(\psi_{AR})$ is in general a mixed state—the partial trace operation need not preserve purity of states. In this case, we call $|\psi\rangle_{AR}$ a *purification* for the mixed state ρ_A with respect to the reference space $\mathcal{H}_R$ [1,2]. Interestingly, it turns out that *any* mixed state can be expressed as the partial trace of a pure state over a larger system:

Theorem 1 (Purifications)**.** *Any mixed state $\rho_A \in \mathcal{H}_A$ has a purification $|\psi\rangle_{AR} \in \mathcal{H}_A \otimes \mathcal{H}_R$ with respect to some reference space $\mathcal{H}_R$, and we may take* $\dim(\mathcal{H}_R) = \dim(\mathcal{H}_A)$.

Proof. Taking a spectral decomposition, we may express ρ_A in the form $\rho_A = \sum_x p_x |x\rangle\langle x|_A$, where the $|x\rangle_A$ are orthonormal. If we take a Hilbert space $\mathcal{H}_R$ isomorphic to $\mathcal{H}_A$, with orthonormal basis $\{|x\rangle_R\}$, then a purification is given by $|\psi\rangle_{AR} = \sum_x \sqrt{p_x}|x\rangle_A \otimes |x\rangle_R$. □

The *canonical purification* of ρ_A is the state $(I_R \otimes \sqrt{\rho_A})|\Gamma\rangle_{RA}$, where I_R is the identity on $\mathcal{H}_R$ and $|\Gamma\rangle_{RA}$ is the "maximally entangled state" $|\Gamma\rangle_{RA} := \sum_i |i\rangle_R|i\rangle_A$ for orthonormal bases $\{|i\rangle_R\}$ and $\{|i\rangle_A\}$ of $\mathcal{H}_R$ and $\mathcal{H}_A$ respectively. Furthermore, one can easily verify that any two purifications are unique up to isometry: if $|\psi\rangle_{AR_1}$ and $|\psi\rangle_{AR_2}$ are purifications of ρ_A, with $\dim(\mathcal{H}_{R_1}) \leq \dim(\mathcal{H}_{R_2})$, then there is an isometry $V : \mathcal{H}_{R_1} \to \mathcal{H}_{R_2}$ $(V^*V = I_{R_1})$ such that $|\psi\rangle_{AR_2} = (I_A \otimes V)|\psi\rangle_{AR_1}$.

A state ρ_A is pure if and only if $\text{rank}(\rho_A) = 1$. In general, $\text{rank}(\rho_A)$ is the smallest dimension of a reference space $\mathcal{H}_R$ over which ρ_A can be purified. This is a consequence of the *Schmidt Decomposition*:

Theorem 2 (Schmidt Decomposition)**.** *Suppose we have a pure bipartite state $|\psi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$. Then we can decompose it as*

$$|\psi_{AB}\rangle = \sum_{i=0}^{d-1} \lambda_i |i\rangle_A |i\rangle_B,$$

where $\{|i\rangle_A\}$ and $\{|i\rangle_B\}$ are orthonormal bases for $\mathcal{H}_A$ and $\mathcal{H}_B$ respectively, and the λ_i are real, strictly positive, and satisfy $\sum_i \lambda_i^2 = 1$. The vector $[\lambda_i]_{i\in\{0,\ldots,d-1\}}$ is called the vector of Schmidt coefficients, and d is called the Schmidt rank, which satisfies

$$d \leq \min\{\dim(\mathcal{H}_A), \dim(\mathcal{H}_B)\}.$$

This is proven by writing $|\psi\rangle_{AB} = \sum_{j=0}^{d_A-1} \sum_{k=0}^{d_B-1} \alpha_{j,k}|j\rangle_A|k\rangle_B$ and considering the singular value decomposition $[\alpha_{j,k}] = \mathbf{U}\mathbf{\Lambda}\mathbf{V}$. The diagonal elements of $\mathbf{\Lambda}$ are the λ_i, and the elements of our orthonormal bases become $|i\rangle_A = \sum_j u_{j,i}|j\rangle_A$ and $|i\rangle_B = \sum_k v_{i,k}|k\rangle_B$, where the $u_{j,i}$ and the $v_{i,k}$ are the entries of $\mathbf{U}$ and $\mathbf{V}$ respectively. The λ_i^2 are the nonzero eigenvalues of $\rho_A = \text{Tr}_{\mathcal{H}_B}(|\psi\rangle\langle\psi|_{AB})$ and $\rho_B = \text{Tr}_{\mathcal{H}_A}(|\psi\rangle\langle\psi|_{AB})$, which both have rank d.

C. Measurement of Quantum States

We have already touched on the idea of measuring or observing a quantum state, but we now formalize this concept. Let $\rho \in \mathcal{H}$ be a quantum state. A noiseless projective quantum measurement of the state ρ corresponds to a complete set of projective operators, $\{\Pi_j\}$, which by definition satisfy the relations $\Pi_i\Pi_j = \delta_{ij}I$ and $\sum_j \Pi_j = I$. Each operator Π_j corresponds to a possible measurement outcome, which we interpret as observing ρ in the space $\mathcal{M}_j$ onto which Π_j projects. The probability of observing the j^{th} outcome is $p_j = \text{Tr}(\Pi_j\rho)$, and observing this outcome immediately transforms ρ to the post-measurement state $\frac{\Pi_j\rho\Pi_j}{\text{Tr}(\Pi_j\rho)}$. If we repeat the measurement $\{\Pi_i\}$ on this new state, we can see that since the Π_i form a complete projective set, we will continue to get outcome j with probability 1.

A quantum measurement can be equivalently expressed as entangling ρ with the state of a measurement device which lives in a Hilbert space of dimension equal to the number of measurement outcomes [3, 4]. The device is initially prepared in the default state $|0\rangle$,

and forms a separable state with ρ in the form $\rho \otimes |0\rangle\langle 0|$. The measurement transforms the state of the device to one of the measurement outcomes, $|j\rangle$, and is equivalent to sending ρ through a channel in the form $\rho \mapsto \sum_j p_j \frac{\Pi_j \rho \Pi_j}{\mathrm{Tr}(\Pi_j \rho)} \otimes |j\rangle\langle j|$.

The most general quantum measurement can be expressed as a Positive Operator Valued Measure (POVM), which is a complete set of nonnegative operators $\Lambda_j \succeq 0$ such that $\sum_j \Lambda_j = I$. This yields the j^{th} measurement outcome with probability $\mathrm{Tr}\,(\Lambda_j \rho)$. From Naimark's Dilation Theorem [5], a POVM applied to a quantum state can be represented by isometrically embedding the state into a higher-dimensional space and performing a projective measurement. Repeated measurement with a POVM will not necessarily yield the same outcome as in the case of a noiseless projective measurement.

III. Quantum Entropy and Quantum Data Compression

A. Quantum Entropy Definitions

The notion of entropy from classical information theory generalizes to the realm of quantum information theory. Given a state $\rho_A \in \mathcal{H}_A$, we define the quantum entropy, or the *von Neumann entropy* (denoted $H(\rho_A)$ or $H(A)_\rho$) as the quantity

$$H(\rho_A) := -\mathrm{Tr}\,(\rho_A \log \rho_A)\,, \tag{3}$$

where $\log \rho_A$ is the base-2 operator logarithm of ρ_A. If ρ_A has a spectral decomposition $\sum_x p_x |x\rangle\langle x|_A$ where $\{|x\rangle_A\}$ is an orthonormal basis for $\mathcal{H}_A$, then $H(\rho_A)$ reduces to the classical entropy of the probability distribution $\{p_x\}$:

$$H(\rho_A) = H(\{p_x\}) := \sum_x -p_x \log p_x. \tag{4}$$

From this observation, it is clear that $H(\rho_A)$ achieves its minimum value of 0 when ρ_A is a pure state, and its maximum value of $\log(\dim(\mathcal{H}_A))$ when ρ_A is the *maximally mixed state* $\pi_A := \frac{1}{d}\sum_x |x\rangle\langle x|_A = \frac{1}{d} I$ (where $d = \dim(\mathcal{H}_A)$ and $\{|x\rangle_A\}$ is an orthonormal basis).

It is important to note several properties of this quantity. The von Neumann entropy is concave in the state ρ_A, and $H(U\rho_A U^*) = H(\rho_A)$ for any isometry U. Furthermore, $H(\rho_A)$ is continuous with respect to the trace distance: $||\rho - \sigma||_1 := \mathrm{Tr}\left(\sqrt{(\rho-\sigma)^*(\rho-\sigma)}\right)$, a result following from the Fannes-Audenaert Inequality (or alternatively, the AFW inequality) [1, 6, 7]. For a pure state $\rho_{AB} = |\phi\rangle\langle\phi|_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$, the marginal entropies of the local operators are equal: $H(\rho_A) = H(\rho_B)$. (This follows from expressing ρ_A and ρ_B in terms of the terms of the components of the Schmidt decomposition of $|\phi\rangle_{AB}$). The entropy is additive over simple product states: $H(\rho_A \otimes \sigma_B) = H(\rho_A) + H(\sigma_B)$. For a *classical-quantum state* of the form $\rho_{XB} = \sum_x p_x |x\rangle\langle x|_X \otimes \rho_B^x$ (where $\{|x\rangle\}$ is an orthonormal basis for $\mathcal{H}_X$), we have

$$H(XB)_\rho = H(X) + \sum_x p_x H(\rho_B^x) \tag{5}$$

where $H(X)$ is the classical entropy of the ensemble $\{x, p_x\}$. Note that classical-quantum states arise when we couple a quantum state ρ (in system B) and a device used to make a classical measurement of it (in system X). The post-measurement state of the two systems will be a classical-quantum state of the above form, where p_x is the probability of observing the x^{th} measurement outcome, and ρ_B^x is the post-measurement state into which ρ_B collapses.

For a bipartite state $\rho_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$, we define the *conditional quantum entropy* as the quantity $H(A|B)_\rho := H(AB)_\rho - H(B)_\rho$. As in the classical case, conditioning reduces entropy: $H(A)_\rho \geq H(A|B)_\rho$. In the quantum case, however, the conditional entropy $H(A|B)_\rho$ can take on negative values. The intuition for this is that in the quantum setting, we can actually be more certain about the entire AB system than just one of its parts. In light of this, we also define the *coherent quantum information*: $I(A\rangle B)_\rho := -H(A|B)_\rho$. If $|\psi\rangle_{ABE} \in \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_E$ is a purification for ρ_{AB}, then using the fact that the local operators ρ_B and ρ_{AE} must have equal entropies, we can verify that $I(A\rangle B)_\rho = H(A|E)_\rho$.

One final quantity of interest is the *quantum relative entropy* between a density operator ρ and a positive semi-definite operator σ over $\mathcal{H}$, defined as

$$D(\rho||\sigma) = \begin{cases} \text{Tr}\left(\rho[\log\rho - \log\sigma]\right) & \text{if } \text{supp}(\rho) \subseteq \text{supp}(\sigma) \\ +\infty & \text{otherwise} \end{cases}. \tag{6}$$

B. Quantum Data Compression

We can formalize the notion of entropy being the information content of a quantum state by considering quantum data compression, also called Schumacher Compression. Suppose we draw pure states from an ensemble $\{p_x, |\psi_x\rangle\}$, forming the product state $|\psi_{x^n}\rangle_{A^n} = |\psi_{x_1}\rangle_{A_1} \otimes \ldots \otimes |\psi_{x^n}\rangle_{A_n}$ with density operator $\rho^{\otimes n}$, where $\rho = \sum_x p_x |\psi_x\rangle\langle\psi_x|$. Let $(|\varphi^\rho\rangle_{RA})^{\otimes n}$ be a purification for $\rho^{\otimes n}$, with density operator $(\varphi^\rho_{RA})^{\otimes n}$. A compression scheme involves an encoding channel $\mathcal{E}_{A^n \to W}$ and a decoding channel $\mathcal{D}_{W \to \hat{A}^n}$ which produces an estimate of the original state. We will define quantum channels formally in Section IV-C, but for now we may take a quantum channel as a map between quantum states of different systems—for example, $\mathcal{E}_{A^n \to W}$ maps quantum states in $\mathcal{H}_{A^n}$ to the intermediate system $\mathcal{H}_W$. If $\mathcal{H}_W$ has size 2^{nR}, we say the scheme compression rate R. It has ϵ-error if $\frac{1}{2}||(\varphi^\rho_{RA})^{\otimes n} - (\mathcal{D}_{W \to \hat{A}^n} \circ \mathcal{E}_{A^n \to W})((\varphi^\rho_{RA})^{\otimes n})||_1 \leq \epsilon$. We call the scheme a (n, R, ϵ) quantum compression code, and say that a rate R is achievable if for all $\delta, \epsilon > 0$, and for n large enough, there is a $(n, R + \delta, \epsilon)$ code.

It turns out that the smallest achievable compression rate is precisely equal to the quantum entropy of the original state:

Theorem 3 (Quantum Data Compression Theorem [8])**.** *If a pure-state quantum information source has density operator ρ_A, the quantum data compression limit (the infimum of achievable rates R) is equal to $H(A)_\rho$.*

This result due to Schumacher [8] can be proven using the quantum analog of typical sets and Shannon-like arguments. The result indicates that a quantum state $(\rho_A)^{\otimes n}$ can be losslessly encoded in a space $\mathcal{H}_W$ of dimension $2^{nH(A)_\rho}$, which can then be transmitted to a receiver with $nH(A)_\rho$ noiseless qubit channel uses. In this sense, we justify the notion of qubits as a fundamental unit of quantum information.

IV. Quantum Information and Quantum Channels

A. Definition of Quantum Mutual Information

Quantum mutual information is defined in terms of the quantum entropies introduced earlier. The *quantum mutual information* is the quantity

$$I(A;B)_\rho := H(A)_\rho + H(B)_\rho - H(AB)_\rho. \tag{7}$$

Similarly, the *conditional quantum mutual information* of a state $\rho_{ABC} \in \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$ is

$$I(A;B|C)_\rho = H(A|C)_\rho + H(B|C)_\rho - H(AB|C)_\rho. \tag{8}$$

The classical notion of Strong Subadditivity still holds for quantum states, which is equivalent to

$$I(A;B|C)_\rho \geq 0. \tag{9}$$

This follows from (and in fact is equivalent to) the non-obvious fact that quantum relative entropy is monotonic with respect to quantum channels [9, 10]. Specifically, given a density operator ρ and a positive semidefinite operator σ over $\mathcal{H}_A$, and a quantum channel $\mathcal{N}$ which sends linear operators from $\mathcal{H}_A$ to $\mathcal{H}_B$, we have the Uhlmann inequality:

$$D(\rho||\sigma) \geq D(\mathcal{N}(\rho)||\mathcal{N}(\sigma)). \tag{10}$$

This monotonicity of quantum relative entropy also implies the quantum data processing inequality: For $\sigma_{A'B'C} = (\mathcal{N}_{A\to A'} \otimes \mathcal{M}_{B\to B'})(\rho_{ABC})$, we have

$$I(A;B|C)_\rho \geq I(A';B'|C)_\sigma. \tag{11}$$

B. Accessible Information and Holevo Information

Suppose we wanted to use a quantum channel to transmit classical information. The information would be encoded in the quantum states which are sent over the channel, and at some point the receiver would have to perform a measurement (by applying a POVM) to obtain a classical message. Suppose the sender encodes a classical message ensemble $\{p_x, x\}$ by transmitting the quantum states $\{\rho_x\}$, leading to the corresponding ensemble of quantum states $\mathcal{E} = \{p_x, \rho_x\}$. The obvious question then becomes how much classical information we can obtain from a POVM. This leads to the notion of *accessible information*:

Definition 2 (Accessible Information)**.** *Suppose we have an ensemble of quantum states $\mathcal{E} = \{p_x, \rho_x\}$ upon which we may perform a POVM $\{\Lambda_y\}$. The accessible information $I_{acc}(\mathcal{E})$ of $\mathcal{E}$ is the maximum mutual information between the classical random variables X and Y, optimized over the choice of POVM: $I_{acc}(\mathcal{E}) = \max_{\{\Lambda_y\}} I(X;Y)$.*

The accessible information can be difficult to compute in general, but we can find a simple upper bound in the *Holevo Information*:

Definition 3 (Holevo Information)**.** *Consider an ensemble of states $\mathcal{E} = \{p_x, \rho_B^x\}$, with the expected density operator $\rho_B = \mathbb{E}_X\{\rho_B^x\} = \sum_x p_x \rho_B^x$. The Holevo information of the ensemble is defined as*

$$\chi(\mathcal{E}) = H(\rho_B) - \sum_x p_x H(\rho_B^x). \tag{12}$$

If we consider the classical-quantum state $\sigma_{XB} = \sum_x p_X(x)|x\rangle\langle x|_X \otimes \rho_B^x$, then the local operator σ_B is precisely the expected density operator ρ_B in the above definition, so we may write the Holevo information succinctly in the form $\chi(\mathcal{E}) = I(X;B)_\sigma$.

Theorem 4 (Holevo Bound [11])**.** *Given an ensemble of quantum states $\mathcal{E} = \{p_x, \rho_x\}$, the Holevo Information is an upper bound on the accessible information: $\chi(\mathcal{E}) \geq I_{acc}(\mathcal{E})$.*

This follows from the quantum data processing inequality. An interesting consequence of this bound is the fact that if the elements of our ensemble ρ_x lie in a Hilbert space of dimension 2^n, we can obtain no more than n bits of classical information from performing a POVM. In particular, if our quantum states are each represented by n qubits, we can gain no more than n bits of classical information from them. This is rather surprising considering that a qubit can take on a continuum of forms (see Sec. IV-D).

C. Definition of a Quantum Channel

We now formally introduce the notion of a quantum channel, which describes the physical transfer and evolution of quantum states and is the basis for using quantum states to convey classical or quantum information. Let $\mathcal{L}(\mathcal{H})$ be the space of linear operators on the Hilbert space $\mathcal{H}$, and $\mathcal{L}(\mathcal{H}_A, \mathcal{H}_B)$ the space of linear maps from $\mathcal{H}_A$ to $\mathcal{H}_B$. We say that a map $\mathcal{N} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ is *trace preserving* if $\mathrm{Tr}\,(\mathcal{N}(X)) = \mathrm{Tr}\,(X)$ for any $X \in \mathcal{L}(\mathcal{H}_A)$. The map $\mathcal{N}$ is *positive* if it takes positive semidefinite operators to positive semidefinite operators, and *completely positive* if it satisfies that $I_R \otimes \mathcal{N} : \mathcal{L}(\mathcal{H}_R \otimes \mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_R \otimes \mathcal{H}_B)$ is a positive map for a reference system R of arbitrary size.

Definition 4. *A quantum channel $\mathcal{N} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ is a linear, completely positive, trace-preserving map, corresponding to a quantum physical evolution.*

To avoid ambiguity in a channel's action, we will sometimes write $\mathcal{N}_{A\to B}$ to emphasize that the channel sends a state in $\mathcal{H}_A$ to one in $\mathcal{H}_B$. Given a bipartite state $\rho_{RA} \in$

$\mathcal{H}_R \otimes \mathcal{H}_A$, we will write $\mathcal{N}_{A\to B}(\rho_{RA})$ as shorthand for $(I_R \otimes \mathcal{N}_{A\to B})(\rho_{RA})$, where I_R indicates the identity map (channel) on linear operators over $\mathcal{H}_R$. The following gives a very useful characterization of quantum channels:

Theorem 5 (Choi-Kraus Theorem)**.** *A map $\mathcal{N} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ is linear, completely positive, and trace-preserving if and only if $\mathcal{N}(X_A) = \sum_{\ell=0}^{d-1} V_\ell X_A V_\ell^*$ where $V_\ell \in \mathcal{L}(\mathcal{H}_A, \mathcal{H}_B)$ and $\sum_{\ell=0}^{d-1} V_\ell^* V_\ell = I_A$. Here, d need not be any larger than $\dim(\mathcal{H}_A) \cdot \dim(\mathcal{H}_B)$.*

The V_ℓ in the above theorem are called *Kraus operators*, and completely determine the action of the quantum channel. The Choi-Kraus Theorem is a direct corollary of Choi's theorem on completely positive maps [12], which states that $\mathcal{N}$ is completely positive if and only if its *Choi operator* (the matrix $[\mathcal{N}(|i\rangle\langle j|_A)]$ with respect to an orthonormal basis $\{|i\rangle_A\}$) is positive semidefinite. The serial concatenation of two channels $\mathcal{N}_{\mathcal{H}_A\to\mathcal{H}_B}$ and $\mathcal{M}_{\mathcal{H}_B\to\mathcal{H}_C}$ with Kraus operators $\{N_k\}$ and $\{M_{k'}\}$ is the channel $\mathcal{M}_{\mathcal{H}_A\to\mathcal{H}_B} \circ \mathcal{N}_{\mathcal{H}_B\to\mathcal{H}_C}$ with Kraus operators $\{M_{k'} \cdot N_k\}$. The parallel concatenation of channels $\mathcal{N}_{\mathcal{H}_A\to\mathcal{H}_B}$ and $\mathcal{M}_{\mathcal{H}_C\to\mathcal{H}_D}$ is the channel $\mathcal{M}_{\mathcal{H}_A\to\mathcal{H}_B} \otimes \mathcal{N}_{\mathcal{H}_B\to\mathcal{H}_C}$ with Kraus operators $\{M_{k'} \otimes N_k\}$.

Schrodinger's Equation tells us that pure states evolve according to unitary processes. It can therefore be useful to note that any quantum channel can be described by a unitary map acting on a purification of a quantum state ρ. This gives rise to an isometry $U : \mathcal{H}_A \to \mathcal{H}_B \otimes \mathcal{H}_R$ such that $\mathrm{Tr}_{\mathcal{H}_R}(UX_AU^*) = \mathcal{N}(X_A)$. Here, $\mathcal{H}_R$ is a reference space of dimension at least the rank of the Choi operator. We call the isometry U an *isometric extension* of $\mathcal{N}$, and sometimes symbolize it as $U^{\mathcal{N}}_{A\to BR}$. One such isometric extension is given by $\sum_j V_j \otimes |j\rangle_R$, where $\{V_j\}$ is a set of Kraus operators for $\mathcal{N}$ and $\{|j\rangle_R\}$ is an orthonormal basis for $\mathcal{H}_R$. If we consider the reference system as starting in a default state $|0\rangle\langle 0|_R$, we can describe $\mathcal{N}$ by a *unitary* extension $U = U^{\mathcal{N}}_{AR\to BR}$ such that $\mathrm{Tr}_{\mathcal{H}_R}(U(X_A \otimes |0\rangle\langle 0|_R)U^*) = \mathcal{N}(X_A)$. Given the channel $\mathcal{N}_{A\to B}$ with isometric extension $U = U^{\mathcal{N}}_{A\to BR}$, we call the corresponding channel $\mathcal{N}^c_{A\to R}(X_A) := \mathrm{Tr}_{\mathcal{H}_B}(UX_AU^*)$ the *complementary* channel to $\mathcal{N}$.

D. A Variety of Quantum Channel Resources

Whereas classical channels are capable of transmitting only classical information, quantum channels can transmit much more, including classical information, quantum information, entanglement, and private information (classical or quantum information to which an eavesdropper can have no access). A *unit resource* of quantum communication is the ability to transmit or consume one unit of classical/quantum information or entanglement. As such, we must quantify these units.

The most familiar of these is classical information, used to describe classical messages. Classical information is measured in the usual classical bits. As we know, a bit takes on a label of 0 or 1, and the information content of a classical message is essentially the number of bits (the length of a binary word) needed to describe the message unambigu-

ously. Quantum information, on the other hand, is quantified in terms of qubits, which are the quantum analogs of bits. A qubit is a pure state in a 2-dimensional Hilbert space $\mathcal{H}_2$. We typically label the elements of an orthonormal basis for $\mathcal{H}_2$ as $|0\rangle$ and $|1\rangle$, so that a qubit is described as a linear combination $\alpha|0\rangle + \beta|1\rangle$ with $|\alpha|^2 + |\beta|^2 = 1$. Similar to the classical case, in which a message is described by a probability distribution on the message alphabet, a quantum message is described by an ensemble of quantum states, which is a mixed state ρ. The quantum information of ρ, loosely speaking, is the number of qubits needed to completely represent it. Just as classical information can be measured by the classical entropy of a message, quantum information can be measured by the quantum entropy of ρ.

The final resource that we wish to quantify is that of entanglement consumption. If a sender Alice and a receiver Bob respectively possess quantum states ρ_A and ρ_B which are entangled (possibly over a great physical distance), this entanglement can be exploited to facilitate the rate at which information can be transferred between them. We quantify entanglement in units of perfectly entangled qubits, which we call ebits. An ebit is a quantum state of the form $\frac{1}{\sqrt{2}}(|00\rangle_{AB} + |11\rangle_{AB})$, where we have used the shorthand $|00\rangle_{AB}$ to denote a quantum state of the form $|0\rangle_A \otimes |0\rangle_B$, and similarly for $|11\rangle_{AB}$. We will discuss the process of converting arbitrary entangled states to ebits in Section VII-A.

Following the convention of [1], we will use the notation $[c \to c]$ to represent a single use of a noiseless classical bit channel. This is a channel that transmits a single bit from the sender Alice to the receiver Bob. Likewise, $[q \to q]$ will denote a single use of a noiseless qubit channel, which is the quantum channel that maps $\alpha|0\rangle_A + \beta|1\rangle_A$ to $\alpha|0\rangle_B + \beta|1\rangle_B$. We will denote the consumption of an ebit by $[qq]$. We will explain what we mean by ebit consumption in the next section. We consider $[c \to c]$, $[q \to q]$ and $[qq]$ to be our "unit resources" of communication because any amount of classical communication, quantum communication, and entanglement consumption can be expressed in terms of multiple uses of noiseless classical bit channels, noiseless qubit channels, and ebits.

V. Quantum Unit Resource Protocols and Unit Resource Inequalities

We can formulate the trade-off between different resources in communication as a kind of algebra. We describe a unit resource communication protocol in the form of an inequality which indicates the rate at which certain resources must be consumed to produce others. For example, an inequality of the form $C[c \to c] + E[qq] \geq Q[q \to q]$, where C, E, and Q are positive real numbers, would indicate a protocol in which C uses of a noiseless classical bit channel and E shared ebits are consumed to simulate Q uses of a noiseless qubit channel. We allow C, Q, and E to assume non-integer values, in which case they correspond to the relative *rates* at which these resources are used. If we allow negative coefficients, we can always express a protocol in the form $0 \geq C[c \to c] + Q[q \to q] + E[qq]$, where a negative coefficient indicates that a resource is consumed and a positive coefficient indicates that a resource is produced or simulated.

We will now describe several fundamental protocols and their corresponding resource inequalities.

A. Entanglement Distribution Protocol

Entanglement distribution [1] is a process by which one use of a noiseless qubit channel is used to produce one shared ebit between sender Alice and receiver Bob. This corresponds to the resource inequality

$$[q \to q] \geq [qq]. \tag{13}$$

As with most protocols, we will assume that any local resources required by Alice and Bob are costless, and let Alice begin with two qubits labeled by A and A' which she prepares in the default state $|0\rangle_A|0\rangle_{A'}$. We also assume that Alice is capable of applying a Hadamard gate operation to a qubit, which is a fundamental quantum operation that transforms the basis of the qubit via the map

$$\begin{aligned} |0\rangle &\mapsto |+\rangle := \frac{|0\rangle + |1\rangle}{\sqrt{2}}, \\ |1\rangle &\mapsto |-\rangle := \frac{|0\rangle - |1\rangle}{\sqrt{2}}. \end{aligned} \tag{14}$$

In general, a Hadamard gate can be performed without actually observing a qubit because it is a linear action defined on the basis elements of our space. This can be important since observing a qubit will collapse it to a single basis element, making it useless for containing quantum information.

In our current situation, however, Alice only applies the Hadamard gate to the basis state $|0\rangle_A$, which induces the transformation $|0\rangle_A|0\rangle_{A'} \mapsto \left(\frac{|0\rangle_A+|1\rangle_A}{\sqrt{2}}\right)|0\rangle_{A'}$. Alice then performs a CNOT gate on the systems A and A'. This quantum gate acts on a pair of qubits (called the "source" and "target" qubits), and induces the map $|a, b\rangle \mapsto |a, a \oplus b\rangle$ where $a, b \in \{0, 1\}$ (a the source and b the target) and $\oplus$ corresponds to mod-2 addition. In this case, using the A system as the source and A' as the target, Alice transforms the state to $\frac{|00\rangle_{AA'}+|11\rangle_{AA'}}{\sqrt{2}}$. Finally, Alice sends the A' qubit to Bob using a single use of a noiseless qubit channel, and resulting in an ebit shared between them in the form $\frac{|00\rangle_{AB}+|11\rangle_{AB}}{\sqrt{2}}$ (where the A system is Alice's share of the ebit, and the B system is Bob's).

B. Super-dense Coding Protocol

The super-dense coding protocol [13] consumes a single use of a noiseless qubit channel and one shared ebit to simulate two uses of a classical bit channel, corresponding to the resource inequality

$$[q \to q] + [qq] \geq 2[c \to c]. \tag{15}$$

This is an interesting use of a quantum channel, since it suggests that if Alice and Bob share stored entanglement upfront, and the cost of using a noiseless qubit channel is the

same as that of a noiseless classical bit channel, then Alice can transmit twice as much classical information to Bob in a given number of uses of the qubit channel as she can in the same number of uses of the bit channel.

In order to understand super-dense coding, we first must define the Pauli operators X and Z, which act on a qubit. X swaps the basis elements $|0\rangle$ and $|1\rangle$, while Z swaps the alternative basis elements $|+\rangle$ and $|-\rangle$. We can express X and Z as matrices with respect to the $\{|0\rangle, |1\rangle\}$ basis as $X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ and $Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. We can apply combinations of these Pauli operators to the first qubit of an ebit to obtain the *Bell basis*, an orthonormal basis for a two-qubit system:

$$\begin{aligned} |\Phi^+\rangle_{AB} &= \frac{1}{\sqrt{2}}(|00\rangle_{AB} + |11\rangle_{AB}) \\ |\Psi^+\rangle_{AB} = X_A|\Phi^+\rangle_{AB} &= \frac{1}{\sqrt{2}}(|01\rangle_{AB} + |10\rangle_{AB}) \\ |\Phi^-\rangle_{AB} = Z_A|\Phi^+\rangle_{AB} &= \frac{1}{\sqrt{2}}(|00\rangle_{AB} - |11\rangle_{AB}) \\ |\Psi^-\rangle_{AB} = Z_AX_A|\Phi^+\rangle_{AB} &= \frac{1}{\sqrt{2}}(|01\rangle_{AB} - |10\rangle_{AB}). \end{aligned} \tag{16}$$

Super-dense coding proceeds as follows: Alice has a 2-bit message (one of four messages), and selects one of the four Pauli operators $W_A \in \{I_A, X_A, Z_A, X_AZ_A\}$ accordingly. She applies the operator to her share of an ebit, $|\Phi^+\rangle_{AB}$, producing a Bell basis element $W_A|\Phi^+\rangle$. She then transfers her share of the ebit (one qubit) to Bob with a single use of a noiseless qubit channel. Bob now has the full state $W_A|\Phi^+\rangle$, and measures it in the Bell basis (an orthogonal projection) to determine Alice's message.

C. Teleportation Protocol

The quantum teleportation protocol [14] is something of a dual to super-dense coding in which the sender utilizes two uses of a classical bit channel along with a single ebit to simulate one use of a noiseless qubit channel:

$$2[c \to c] + [qq] \geq [q \to q]. \tag{17}$$

In teleportation, Alice wants to send a qubit $|\psi\rangle_{A'}$ to Bob. Alice and Bob share an ebit $|\Phi^+\rangle_{AB}$. With some simple algebra, we can re-express the product state $|\psi\rangle_{A'}|\Phi^+\rangle_{AB}$ in the equivalent form

$$\frac{1}{2}(|\Phi^+\rangle_{A'A}|\psi\rangle_B + |\Phi^-\rangle_{A'A}Z_B|\psi\rangle_B + |\Psi^+\rangle_{A'A}X_B|\psi\rangle_B + |\Psi^-\rangle_{A'A}X_BZ_B|\psi\rangle_B), \tag{18}$$

where X_B, Z_B are Pauli operators acting on Bob's state.

Alice performs a measurement on her two-qubit state $A'A$ in the Bell basis, and sends a 2-bit classical message to Bob to tell him which of the four outcomes she observes. Bob then knows his state is in the form $|\psi\rangle_B$, $Z_B|\psi\rangle_B$, $X_B|\psi\rangle_B$ or $X_BZ_B|\psi\rangle_B$ depending on

what Alice tells him. By applying the appropriate inverse Pauli operator, Bob puts his qubit in the state $|\psi\rangle_B$, effectively "teleporting" Alice's original qubit to Bob (note that Alice's original qubit collapsed during her Bell basis measurement, so that the quantum "no-cloning" theorem is not violated).

D. The Unit Resource Capacity Region

It is natural to ask whether we can completely characterize the trade-off region of the three unit resources $[c \to c]$, $[q \to q]$, and $[qq]$. In other words, if we express all unit resource inequalities in the form $0 \geq C[c \to c] + Q[q \to q] + E[qq]$, what is the set of all triples (C, Q, E) corresponding to achievable protocols involving only the unit resources? We call the closure of this set the *Unit Resource Capacity Region*, denoted C_U.

The three protocols we have considered—entanglement distribution, super-dense coding, and teleportation—correspond to the ordered triples $(0, -1, 1)$, $(2, -1, -1)$, and $(-2, 1, -1)$ respectively. Timesharing between these protocols, we can achieve any point in the convex cone of these three ordered triples. It turns out that this yields the entire Unit Resource Capacity Region:

Theorem 6 (Unit Resource Capacity Region [15])**.** *The Unit Resource Capacity Region, C_U, is precisely the convex cone of the (C, Q, E) triples $(0, -1, 1)$, $(2, -1, -1)$, and $(-2, 1, -1)$ corresponding to entanglement distribution, super-dense coding, and teleportation respectively.*

Shortly, we will discuss communication protocols which involve the use of an arbitrary quantum channel $\mathcal{N}$. We will see that when we allow ourselves the additional resource of uses of $\mathcal{N}$, we may be able to produce values of C, Q, and E which fall outside the region C_U.

VI. Four Different Quantum Channel Capacities

We are now ready to discuss the rate at which we can reliably communicate information over a quantum channel. Whereas in the classical case we have a single notion of channel capacity (the maximum rate at which we can communicate classical bits reliably), a quantum channel has *four* different capacities associated with it:

1. **Classical Capacity $C(\mathcal{N})$:** The best rate at which a sender can transmit classical information (bits) over the channel.
2. **Entanglement-Assisted Classical Capacity $C_E(\mathcal{N})$:** The best rate at which a sender can transmit classical information when the sender and receiver share an arbitrary number of quantum states (ebits).
3. **Private Classical Capacity $P(\mathcal{N})$:** The best rate for sending classical informa-

tion to achieve high fidelity between sender and receiver *without* leaking information to the environment.

4. **Quantum Capacity $Q(\mathcal{N})$:** The best rate for sending quantum information (qubits) over the channel.

Unlike classical channels, for which classical Shannon theory completely describes the channel capacity in terms of the maximum mutual information between sent and received messages, the four quantum channel capacities remain to be fully characterized. In fact, only the entanglement-assisted capacity $C_E(\mathcal{N})$ has been successfully expressed in a simple form for general channels. Much has been learned about all four capacities, however, using the quantum analog of Shannon theory. An excellent reference on "Quantum Shannon Theory" is provided by Wilde's book [1].

We must emphasize that the results in this section typically assume that our sender Alice and receiver Bob have access to any theoretically realizable quantum resources. In particular, Alice can transmit information encoded in arbitrary quantum states, and Bob can apply an arbitrary POVM to measure the channel output. As such, these capacities will be upper bounds the maximal rates of communication that can be achieved using states and receivers which can be feasibly implemented in the lab. In Section VIII-C we will describe some of the capacities of free-space optical communication when restricted to using common optical states. These are different from the ultimate capacity of the free-space optical channel, described in Section VIII-B.

A. Classical Capacity

Let us formally define the classical capacity of a quantum channel, for which we must first describe the classical communication information processing task. The sender Alice has a message $m \in \{1, ..., |\mathcal{M}|\}$ represented by random variable M. She prepares a state $\rho^m_{A'^n}$ which she sends to Bob over n parallel uses of the channel $\mathcal{N}$. Bob receives the state $\mathcal{N}^{\otimes n}(\rho^m_{A'^n})$, which he measures with a POVM $\{\Lambda_m\}$. He estimates Alice's original message as a random variable M'. The probability of a correct estimate is $P(M' = m|M = m) = \mathrm{Tr}\left(\Lambda_m \mathcal{N}^{\otimes n}(\rho^m_{A'^n})\right)$ (not to be confused with the above notation for the Private Classical Capacity of a channel), and the error probability for message m is $p_e(m) = 1 - P(M' = m|M = m) = \mathrm{Tr}\left((I - \Lambda_m)\mathcal{N}^{\otimes n}(\rho^m_{A'^n})\right)$. We set $p^*_e = \max_m p_e(m)$ and say that the code has error less than $\epsilon \in [0, 1]$ if $p^*_e \leq \epsilon$. The classical communication rate is $C := \frac{1}{n} \log |\mathcal{M}|$, and we call this an (n, C, ϵ) code. We say that a rate C is *achievable* for $\mathcal{N}$ if there is an $(n, C - \delta, \epsilon)$ code for all $\delta > 0$, $\epsilon \in [0, 1]$, and large enough n. Expressed in the language of resource inequalities, the classical capacity of $\mathcal{N}$ is the largest value of C such that there is an achievable protocol corresponding to $\langle \mathcal{N} \rangle \geq C[c \to c]$, where $\langle \mathcal{N} \rangle$ represents the resource of a single use of the noisy quantum channel $\mathcal{N}$.

Holevo [16] and (separately) Schumacher and Westmoreland [17] proved that the classical capacity can be described in terms of the Holevo information of $\mathcal{N}$:

Definition 5 (Holevo Information of a Channel)**.** *Suppose Alice draws states from an ensemble* $\{p_x, \rho_A^x\}$*, with the corresponding classical-quantum state* $\rho_{XA} = \sum_x p_x |x\rangle\langle x| \otimes \rho_A^x$*. She sends the A subsystem through the quantum channel* $\mathcal{N}_{A\to B}$*, yielding the state* $\rho_{XB} = \sum_x p_x |x\rangle\langle x| \otimes \mathcal{N}_{A\to B}(\rho_A^x)$*. If we define the states* $\sigma_B^x := \mathcal{N}_{\mathcal{A}\to\mathcal{B}}(\rho_{\mathcal{A}}^{\S})$*, then the Holevo Information of* $\mathcal{N}$ *is*

$$\chi(\mathcal{N}) := \max_{\rho_{XA}} I(X;B)_\rho \tag{19}$$

$$= \max_{\{p_x, \rho_A^x\}} \left[H\left(\sum_x p_x \sigma_B^x\right) - \sum_x p_x H(\sigma_B^x) \right]. \tag{20}$$

The Classical Capacity Theorem (also called the "HSW" Theorem after Holevo, Schumacher, and Westmoreland [16, 17]) shows that by measuring all the channel outputs collectively (i.e. exploiting entanglement of the output states), we can achieve the Holevo information as a rate of classical communication.

Theorem 7 (Classical Capacity (HSW) Theorem)**.** *The Holevo information is an achievable rate of classical communication over the quantum channel* $\mathcal{N}$*:* $C(\mathcal{N}) \geq \chi(\mathcal{N})$*. In fact,* $C(\mathcal{N})$ *is equal to the regularization of the Holevo information:*

$$C(\mathcal{N}) = \chi_{reg}(\mathcal{N}) := \lim_{n\to\infty} \frac{1}{n} \chi(\mathcal{N}^{\otimes n}). \tag{21}$$

The achievability of $\chi(\mathcal{N})$ as a rate of classical communication can be proven using what is known as the Packing Lemma [1, 18], a useful tool for showing the existence of a set of quantum states and a corresponding POVM which yield low estimation error. Essentially, given an ensemble of states $\{p_x, \rho_A^x\}$ which achieves the Holevo information in Definition 7, sequences of states drawn from this ensemble can be concatenated into product states which serve as codewords. The argument follows a Shannon-style approach [1].

For a long time, it was an open question whether $\chi(\mathcal{N})$ was additive over parallel concatenations of channels. That is, does $\chi(\mathcal{M} \otimes \mathcal{N}) = \chi(\mathcal{M}) + \chi(\mathcal{N})$? If so, this would imply that the classical capacity can be expressed as the "single-letter" form $C(\mathcal{N}) = \chi(\mathcal{N})$. Note that for classical channels, this is indeed the case: the maximal mutual information between input and output messages is additive over parallel uses of two channels, which is why classical capacity over classical channels has a single-letter form. For quantum channels, however, this turns out to be false. Hastings [19] demonstrated the existence of channels $\mathcal{M}$ and $\mathcal{N}$ for which the Holevo information is strictly superadditive: $\chi(\mathcal{M} \otimes \mathcal{N}) > \chi(\mathcal{M}) + \chi(\mathcal{N})$. There are, however, particular examples (the so-called "entanglement-breaking" channels [20]) for which the Holevo information is additive over parallel channel uses, so for these we can indeed express the classical capacity as $\chi(\mathcal{N})$.

A further and perhaps more important question is whether the classical capacity $C(\mathcal{N})$ is *itself* additive over channels. $C(\mathcal{N})$ is additive over multiple copies of the *same*

channel, since it follows from the regularization expression that

$$C(\mathcal{N}^{\otimes m}) = \lim_{n\to\infty} \frac{1}{n}\chi(\mathcal{N}^{\otimes mn}) = m \cdot \left(\lim_{n\to\infty} \frac{1}{mn}\chi(\mathcal{N}^{\otimes mn})\right) = m \cdot C(\mathcal{N}). \tag{22}$$

But for different channels $\mathcal{M}$ and $\mathcal{N}$, it is still unknown whether $C(\mathcal{M}\otimes\mathcal{N}) = C(\mathcal{M}) + C(\mathcal{N})$. Note that this is not prohibited by the nonadditivity of $\chi(\mathcal{N})$.

B. Entanglement-Assisted Classical Communication

It stands to reason that we can boost the classical capacity of a quantum channel by exploiting shared entanglement between sender and receiver. Holevo's bound shows us that using a noiseless qubit channel (through which Alice can transmit one qubit per channel use to Bob), Bob can extract at most one bit of classical information for each qubit that Alice sends. The super-dense coding protocol, however, shows that if Alice and Bob share one ebit of entanglement, Bob can receive *two* bits per sent qubit. How does this generalize to arbitrary channels?

The answer was provided by Bennett, Shor, Smolin, and Thapliyal [21] through the Entanglement-Assisted Classical (EAC) Capacity Theorem, which characterizes the highest achievable rate C of classical communication in the resource inequality $\langle\mathcal{N}\rangle + \infty[qq] \geq C[c\to c]$. This resource inequality asks for the number of transmissible bits per channel use given infinite shared ebits.

First we describe the entanglement-assisted classical communication protocol and formally define the entanglement-assisted classical capacity. Alice and Bob share copies of a pure entangled state $\Psi_{T_AT_B}$. Alice selects a message $m \in \mathcal{M}$ and uses an encoding channel $\mathcal{E}^m_{T_A\to A'^n}$ on her share of $\Psi_{T_AT_B}$. She then sends the A'^n system to Bob using n independent uses of the noisy channel, producing the state $\mathcal{N}_{A'^n\to B^n}(\mathcal{E}^m_{T_A\to A'^n}(\Psi_{T_AT_B}))$, where $\mathcal{N}_{A'^n\to B^n} = (\mathcal{N}_{A'\to B})^{\otimes n}$. Bob then measures his composite system B^nT_B with a POVM $\{\Lambda^m_{B^nT_B}\}$ to produce an estimate m' for m. Similar to the case of ordinary classical communication, the probability of error for message m is $p_e(m) = \operatorname{Tr}\left((I - \Lambda^m_{B^nT_B})\mathcal{N}_{A'^n\to B^n}(\mathcal{E}^m_{T_A\to A'^n}(\Psi_{T_AT_B}))\right)$, and we set $p^*_e = \max_{m\in\mathcal{M}} p_e(m)$. The rate of communication is $C = \frac{1}{n}\log|\mathcal{M}|$, and if $p^*_e \leq \epsilon$ we say this is an (n, C, ϵ) entanglement-assisted classical code. A rate C is achievable if there is a $(n, C-\delta, \epsilon)$ entanglement-assisted classical code $\forall\epsilon\in(0,1)$, $\delta > 0$, and n large enough.

The Entanglement-Assisted Classical Capacity Theorem describes the highest rate of classical communication given an arbitrary amount of shared entanglement between Alice and Bob, which turns out to be equal to the *mutual information of the quantum channel*:

Definition 6 (Mutual Information of a Quantum Channel)**.** *The mutual information of a quantum channel $\mathcal{N}_{A'\to B}$ is the quantity $I(\mathcal{N}) := \max_{\rho_{AA'}} I(A;B)_{\mathcal{N}(\rho)}$, where $\rho_{AB} = \mathcal{N}_{A'\to B}(\rho_{AA'})$. In fact, by purifying $\rho_{AA'}$ and exploiting the quantum data-processing inequality, it can be shown that it is enough to optimize over pure states*

$\varphi_{AA'}$. *Thus we may formally define* $I(\mathcal{N})$ *as*

$$I(\mathcal{N}) := \max_{\{\varphi_{AA'} \ pure\}} I(A;B)_{\mathcal{N}(\varphi)}. \tag{23}$$

$I(\mathcal{N}_{A\to B})$ is a measure of the ability of the channel to preserve quantum correlations between A and B. It turns out that, unlike Holevo information, the quantum channel mutual information obeys additivity: $I(\mathcal{M}\otimes\mathcal{N}) = I(\mathcal{M}) + I(\mathcal{N})$, a fact which follows from strong subadditivity. This is what allows us to find a single-letter expression for $C_E(\mathcal{N})$.

Theorem 8 (Entanglement-Assisted Capacity (EAC) [21])**.** *The entanglement-assisted classical capacity is*

$$C_E(\mathcal{N}) = I(\mathcal{N}) = \max_{\varphi_{AA'}} I(A;B)_\rho, \tag{24}$$

where $\varphi_{AA'}$ *is a pure bipartite state and* $\rho_{AB} = \mathcal{N}_{A'\to B}(\varphi_{AA'})$*. Furthermore, for any such states,* $\langle\mathcal{N}\rangle + H(A)_\rho[qq] \geq I(A;B)_\rho[c\to c]$ *is an achievable protocol, so* $C_E(\mathcal{N})$ *can be achieved with an entanglement consumption rate of* $H(A)_\rho$ *ebits per channel use for the corresponding* ρ_{AB}*.*

We will omit the details of the proof, except to mention that the achievability of such a protocol can be shown by using a random coding argument and exploiting the aforementioned Packing Lemma to prove the existence of an accurate decoding POVM [1]. Note that due to the single-letter additivity of $I(\mathcal{N})$, the entanglement-assisted capacity $C_E(\mathcal{N})$ is itself additive.

C. Private Classical Communication

A unique feature of quantum channels is the ability to transmit classical information which is immune to eavesdropping, in the sense that the quantum mutual information between the sent quantum states and an eavesdropper is small. The "eavesdropper" can also be viewed as the environment of the quantum channel, which itself can interact with the transmitted quantum states, inadvertently observing and transforming them.

In private classical communication, Alice selects a message $m \in \mathcal{M}$, prepares a state $\rho^m_{A'^n}$, and sends to Bob $\mathcal{N}_{A'^n\to B^n}(\rho^m_{A'^n})$ over n uses of the quantum channel $\mathcal{N}_{\mathcal{A}'\to\mathcal{B}}$ (where $\mathcal{N}_{A'^n\to B^n} := (\mathcal{N}_{\mathcal{A}'\to\mathcal{B}})^{\otimes n}$). Bob then measures the received state using a POVM $\{\Lambda_m\}$ with error probability $p_e(m) = \mathrm{Tr}\left((I-\Lambda_m)\mathcal{N}_{A'^n\to B^n}(\rho^m_{A'^n})\right)$, as before, and likewise the rate of the code is $P = \frac{1}{n}\log|\mathcal{M}|$. But now, an eavesdropper Eve has access to a channel $\hat{\mathcal{N}}_{A'\to E}(\sigma) = \mathrm{Tr}_B\left(U\sigma U^\dagger\right)$, where $U = U^{\mathcal{N}}_{A'\to BE}$ is an isometric extension of $\mathcal{N}_{A'\to B}$ (in keeping with the interpretation of Eve as the environment of the channel).

Define $p^*_e := \max p_e(m)$ and $\omega^m_{E^n} := \hat{\mathcal{N}}_{A'^n\to E^n}(\rho^m_{A'^n})$. Then for $\epsilon > 0$, we say that this communication scheme is an (n, P, ϵ) code if $p^*_e \leq \epsilon$ and $\frac{1}{2}||\omega^m_{E^n} - \sigma_{E^n}||_1 \leq \epsilon$ for all m, where σ_{E^n} is a constant state. For ϵ small, the condition that $\frac{1}{2}||\omega^m_{E^n} - \sigma_{E^n}||_1 \leq \epsilon$

intuitively means that Eve cannot distinguish between the different states she intercepts. Using the AFW inequality, it can be shown formally that this implies $I(M;E^n)_\omega$ is small, so that Eve gains little information about Alice's message.

In order to describe the private classical capacity in terms of resource inequalities, we would have to introduce a new unit resource, that of a use of a "private" bit channel, $[c \to c]_{\text{priv}}$, in which case $P(\mathcal{N})$ can be defined as the largest rate P such that $\langle \mathcal{N} \rangle \geq P[c \to c]_{\text{priv}}$. We chose not to include this in our discussion of the unit resource capacity region in Section V for simplicity.

To find the maximum rate of private classical communication, we must define the *private information* of a quantum channel:

Definition 7 (Private Information of a Quantum Channel)**.** *Let $\mathcal{N}_{A' \to \mathcal{B}}$ be a quantum channel with an isometric extension $U = U^{\mathcal{N}}_{A' \to BE}$. The private information of $\mathcal{N}$ is the quantity*

$$P^{(1)}(\mathcal{N}) := \max_{\rho_{XA'}} \left[I(X;B)_\sigma - I(X;E)_\sigma \right], \tag{25}$$

where the maximization is taken over classical-quantum states $\rho_{XA'} = \sum_x p_x |x\rangle\langle x|_X \otimes \rho^x_{A'}$, and $\sigma_{XBE} = \mathcal{U}^{\mathcal{N}}_{A' \to BE}(\rho_{XA'}) := \sum_x p_x |x\rangle\langle x|_X \otimes U \rho^x_{A'} U^\dagger$.

The private information of $\mathcal{N}_{A' \to B}$ is a measure of the classical correlations that Alice (A') can send to Bob (B) minus the classical correlations that are leaked to Eve (E, the environment). $P(\mathcal{N})$ is a nonnegative quantity due to symmetry between B and E. It was Devetak [22] and Cai-Winter-Young [23] who described the private classical capacity $P(\mathcal{N})$ in terms of the private information:

Theorem 9 (Private Classical Capacity)**.** *Let $\mathcal{N}$ be a quantum channel. The private information $P^{(1)}(\mathcal{N})$ is an achievable rate of the private classical capacity: $P(\mathcal{N}) \geq P^{(1)}(\mathcal{N})$. In fact, the private classical capacity is equal to the regularization of the private information:*

$$P(\mathcal{N}) = P^{(1)}_{reg}(\mathcal{N}) := \lim_{k \to \infty} \frac{1}{k} P^{(1)}(\mathcal{N}^{\otimes k}). \tag{26}$$

The achievability of $P(\mathcal{N})$ can be proven with a random coding argument. If the sender Alice randomly selects a large enough set of codeword states, then the Packing Lemma guarantees a POVM for receiver Bob to decode Alice's message. Furthermore, the "Covering Lemma" [1], a corollary of the operator Chernoff bound from [24], shows the existence of a "fake" expected density operator which Eve cannot distinguish from Alice's true message state density operator.

For an arbitrary channel $\mathcal{N} = \mathcal{N}_{A \to B}$, neither $P^{(1)}(\mathcal{N})$ nor $P(\mathcal{N})$ are additive in general [25, 26]. One exception is in the case of *degradable* channels, defined by the property that the complementary channel $\mathcal{N}^c_{A \to E}$ can be expressed as a series of channels $\mathcal{D}_{B \to E} \circ \mathcal{N}_{A \to B}$. For degradable channels, the private information is additive, hence equal to its regularization and to the private classical capacity [27].

D. Quantum Capacity

In quantum communication, Alice begins with some shared entanglement with a reference system R in the form of some pure state φ_{RA}. She wishes to transfer the correlation between her system and R to Bob, to produce a state ω_{RB} which is close to φ_{RA} in trace distance. To this end, Alice uses an encoding channel $\mathcal{E}_{A\to A'^n}$ to prepare her state for transmission to Bob via n uses of the quantum channel $\mathcal{N}_{A'\to B'}$ (forming the channel $\mathcal{N}_{A'^n\to B'^n} = (\mathcal{N}_{A'\to B'})^{\otimes n}$). Bob receives the state $\mathcal{N}_{A'^n\to B'^n}(\mathcal{E}_{A\to A'^n}(\varphi_{RA}))$, which he decodes with a channel $\mathcal{D}_{B'^n\to B}$ to get $\omega_{RB} := \mathcal{D}_{B'^n\to B}(\mathcal{N}_{A'^n\to B'^n}(\mathcal{E}_{A\to A'^n}(\varphi_{RA})))$. The rate of this code is defined as $Q = \frac{1}{n}\log\dim(\mathcal{H}_A)$. If $\frac{1}{2}||\varphi_{RA} = \omega_{RA}||_1 \leq \epsilon$, and we call it an (n, Q, ϵ) code. The quantum capacity $Q(\mathcal{N})$ is the supremum of all Q for which there exists an (n, Q, ϵ) code for any $\epsilon > 0$ and some n. It is the maximum Q for which there is a protocol that achieves the resource inequality $\langle\mathcal{N}\rangle \geq Q[q\to q]$.

We can express a channel's quantum capacity in terms of its *coherent information*:

Definition 8 (Coherent Information of a Channel)**.** *The coherent information $Q^{(1)}(\mathcal{N})$ of a channel $\mathcal{N} = \mathcal{N}_{A'\to B}$ is the quantity*

$$Q^{(1)}(\mathcal{N}) := \max_{\{\varphi_{AA'}\ pure\ \}} I(A\rangle B)_\rho, \tag{27}$$

where $\rho_{AB} = \mathcal{N}_{A'\to B}(\varphi_{AA'})$.

If we take a unitary extension $U = U^{\mathcal{N}}_{A'\to BE}$ for the channel, and set $\sigma_{ABE} = U(\varphi_{AA'}\otimes|0\rangle\langle 0|_E)U^\dagger$. Recall the fact that the marginal entropies of the associated local operators are equal. Then we can write $I(A\rangle B)_\sigma = H(B)_\sigma - H(AB)_\sigma = H(B)_\sigma - H(E)_\sigma$. This allows us to express the coherent information of the channel as

$$Q^{(1)}(\mathcal{N}) = \max_{\{\varphi_{AA'}\ \text{pure}\ \}} H(B)_\sigma - H(E)_\sigma, \tag{28}$$

where σ_{ABE} is defined as above.

The coherent information is nonnegative, and with a little work, it can be shown that it is upper-bounded by the private information: $Q^{(1)}(\mathcal{N}) \leq P^{(1)}(\mathcal{N})$. Lloyd [28], Shor [29], and Devetek [22] (separately) were able to prove that $Q^{(1)}(\mathcal{N})$ is an achievable rate for quantum communication:

Theorem 10 (Quantum Communication)**.** *The coherent information $Q^{(1)}(\mathcal{N})$ is an achievable rate for quantum communication over a quantum channel $\mathcal{N}$, so we have $Q(\mathcal{N}) \geq Q^{(1)}(\mathcal{N})$. In fact, the quantum capacity is equal to the regularization of the channel's coherent information:*

$$Q(\mathcal{N}) = \lim_{k\to\infty}\frac{1}{k}Q^{(1)}(\mathcal{N}^{\otimes k}). \tag{29}$$

Neither the coherent information $Q^{(1)}(\mathcal{N})$ [30] nor the quantum capacity $Q(\mathcal{N})$ [31] is additive in general. One exception is in the case of degradable channels [32], for

which it turns out that $Q^{(1)}(\mathcal{N}) = P^{(1)}(\mathcal{N})$ [27]. It is also additive for channels with the "positive partial transpose" (PPT) property [33, 34]. These are channels which send bipartite input states to bipartite output states whose local operators are positive semidefinite. In fact, for PPT channels, it can actually be shown that $Q^{(1)}(\mathcal{N}) = 0$.

VII. Some Achievable Quantum Resource Trade-offs

A. Entanglement Manipulation

Up to this point, we have quantified entanglement in terms of ebits. But what if a sender Alice and receiver Bob each have a share of copies of some other pure bipartite state, ψ_{AB}? It turns out that Alice and Bob can convert copies of ψ_{AB} to a certain number of copies of any other entangled pure state ϕ_{AB}, a process called *entanglement manipulation*. Here, Alice and Bob try to convert n copies of ψ_{AB} to m copies of ϕ_{AB} using an LOCC channel $\Lambda^{(n)}_{A^nB^n \to A^mB^m}$. "LOCC" stands for "local operations and classical communication," and indicates that the channel must be enacted by Alice and Bob separately performing local quantum measurements with classical outputs which they may communicate to each other. If we set $\omega_{A^mB^m} := \Lambda^{(n)}_{A^nB^n \to A^mB^m}(\psi^{\otimes n}_{AB})$, then the protocol has ϵ-error if $\frac{1}{2}||\omega_{A^mB^m} - \phi^{\otimes m}_{AB}||_1 \leq \epsilon$. We call this an (n, E, ϵ) protocol with rate $E = m/n$. A rate E is achievable if for all $\delta, \epsilon > 0$, and large enough n, there exists an (n, E, ϵ) protocol for entanglement manipulation, and we define the entanglement manipulation limit $E(\psi \to \phi)$ to be the supremum of all achievable rates.

This limit turns out to be equal to the ratio of the entropies of the two states:

Theorem 11 (Entanglement Manipulation)**.** *The entanglement manipulation limit for* $\psi_{AB} \to \phi_{AB}$ *is* $E(\psi \to \phi) = \frac{H(A)_\psi}{H(A)_\phi}$.

The proof relies on achieving this limit using a two-part process. The first step is to perform "entanglement concentration" [35] to convert n copies of ψ_{AB} to approximately $nH(A)_\psi$ ebits. The second is to perform "entanglement dilution," a process which converts these ebits to copies of ϕ_{AB} at a rate of $H(A)_\phi$ ebits per copy [36,37]. It should be noted that entanglement dilution requires some amount of classical communication between Alice and Bob, but this amount is negligible (having a bit rate sublinear in n) [38, 39]. The entanglement manipulation theorem justifies the idea of using ebits to quantify entanglement. It establishes that the number of ebits of entanglement associated to a state ψ_{AB} is approximately $H(A)_\psi$, and these can be extracted via entanglement concentration.

B. Entanglement-Assisted Quantum Communication

There are several other useful protocols worth noting. The first can be proven in a similar fashion as the entanglement-assisted classical capacity theorem, and involves exploiting the *coherent communication identity* [1, 40]. We will omit a discussion of coherent quantum communication, but [1] provides a good reference on the subject.

The result reveals an achievable protocol for communicating qubits by consuming ebits:

Theorem 12 (Entanglement-Assisted Quantum Communication)**.** *Given a quantum channel $\mathcal{N} = \mathcal{N}_{A' \to B}$ with isometric extension $U = U^{\mathcal{N}}_{A' \to BE}$, and a pure state $|\varphi\rangle_{A'A}$, the following is an achievable protocol:*

$$\langle \mathcal{N} \rangle + \frac{1}{2} I(A;E)_{\varphi}[qq] \geq \frac{1}{2} I(A;B)_{\varphi}[q \to q], \tag{30}$$

where $|\varphi\rangle_{ABE} := U^{\mathcal{N}}_{A' \to BE} |\varphi\rangle_{A'A}$. That is, asymptotically, for each channel use we can consume $\frac{1}{2} I(A;E)_{\varphi}$ ebits and communicate $\frac{1}{2} I(A;B)_{\varphi}$ qubits.

An alternative protocol for entanglement-assisted quantum communication could be derived, for instance, by starting with the entanglement-assisted classical communication protocol to consume ebits to communicate classical bits, and using teleportation to use these bits to transmit qubits (at the cost of further ebit consumption). It can be verified, however, that this would require a greater rate of ebit consumption than that of Theorem 12.

It is also worth noting that the achievability of the coherent information $I(A\rangle B)$ as a rate of quantum communication (Theorem 10) can be realized as a corollary of Theorem 12 by combining the above protocol with entanglement distribution at a rate of $\frac{1}{2} I(A;E)_{\rho}$ and noting that $I(A\rangle B)$ can be equivalently expressed as $\frac{1}{2}[I(A;B) - I(A;E)]$.

C. Noisy Super-dense Coding

Another protocol of note is *noisy super-dense coding*.

Theorem 13 (Noisy Super-Dense Coding)**.** *Suppose a sender Alice and a receiver Bob share copies of a state $\rho_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$. There is an achievable protocol for quantum-assisted classical communication with a shared quantum state with inequality:*

$$\langle \rho_{AB} \rangle + H(A)_{\rho}[q \to q] \geq I(A;B)_{\rho}[c \to c]. \tag{31}$$

Here, $\langle \rho_{AB} \rangle$ indicates consumption of a copy of ρ_{AB}.

When the state ρ_{AB} is a perfect ebit, this reduces to the usual super-dense coding. The entanglement-assisted classical capacity theorem generalized super-dense coding, using an arbitrary (noisy) quantum channel instead of a noiseless qubit channel. Noisy super-dense coding is another generalization, which assumes entanglement in the form of a noisy shared state ρ_{AB} as opposed to perfect ebits. Theorems 12 and 13 are corollaries of the coherent communication identity and the results in [41] and [42].

D. Trade-Off Coding

The protocol from the entanglement-assisted classical capacity theorem reveals that given a channel $\mathcal{N}_{A' \to B}$, ebits can be consumed at a rate of $H(A)_{\rho}$ per channel use

in order to convey the optimal $I(A;B)_\rho$ bits of classical information, where $\rho_{AB} = \mathcal{N}_{A'\to B}(\varphi_{AA'})$. Since maintaining and distributing entanglement can be costly, a natural question to ask is at what rate classical information can be sent while consuming ebits at a lower rate? Shor and Devetak [32, 43] attempted to answer this question in works which were generalized by Hsieh and Wilde [44], who developed a technique aptly called *trade-off coding*.

Theorem 14 (CE Trade-off Coding)**.** *The following corresponds to an achievable protocol for entanglement-assisted classical communication over a channel $\mathcal{N}_{\mathcal{A}'\to\mathcal{B}}$:*

$$\langle \mathcal{N} \rangle + H(A|X)_\rho[qq] \geq I(AX;B)_\rho[c \to c], \tag{32}$$

where $\rho_{XAB} = \sum_x p_x |x\rangle\langle x|_X \otimes \mathcal{N}_{A'\to B}(\varphi^x_{AA'})$, with the $\varphi^x_{AA'}$ pure.

The abbreviation "CE" in "CE Trade-off Coding" indicates that the theorem addresses a trade-off between the rate C of classical communication and the rate E of entanglement consumption. When X is endowed with the trivial ensemble $p_0 = 1$, we can see that the above protocol collapses to that of entanglement-assisted classical communication from Theorem 8. Theorem 14 can be proven by constructing a protocol in which Alice starts with a high-dimensional quantum state and uses many entanglement-assisted classical (EAC) codes to encode different parts of this state independently. Alice encodes into the *entire* state a message to Bob which indicates how to find each of these EAC codes. The Packing Lemma can be applied to guarantee the existence of such a scheme, as well as a POVM which Bob can use to detect Alice's message. Trade-off coding can achieve rates of classical communication with lower rates of ebit consumption than time-sharing between entanglement-assisted classical communication and regular (non-entanglement-assisted) classical communication (from the HSW Theorem). In fact, time-sharing can actually be realized as a special case of trade-off coding.

As a corollary of CE Trade-off Coding, we can derive the following family of achievable protocols for trading off between rates of classical communication (C), quantum communication (Q) and entanglement consumption (E) [44]:

Theorem 15 (CQE Trade-off Coding)**.** *Let $\mathcal{N} = \mathcal{N}_{A'\to B}$ be a quantum channel with isometric extension $U = U^{\mathcal{N}}_{A'\to BE}$. The following is an achievable protocol:*

$$\langle \mathcal{N} \rangle + \frac{1}{2} I(A;E|X)_\rho[qq] \geq \frac{1}{2} I(A;B|X)_\rho[q \to q] + I(X;B)_\rho[c \to c], \tag{33}$$

where $\rho_{XABE} := \sum_x p_X(x) |x\rangle\langle x|_X \otimes U \varphi^x_{AA'} U^\dagger$ and the $\varphi^x_{AA'}$ are pure.

The CQE Trade-off Coding protocol is a particularly important result, for when combined with teleportation, super-dense coding and entanglement distribution, it is sufficient to achieve any task in dynamic quantum Shannon theory involving the noisy channel $\mathcal{N}$ and the three unit resources, as shown by Hsieh and Wilde [15, 44, 45]:

Theorem 16 (Dynamic Capacity Region)**.** *Given a quantum channel $\mathcal{N} = \mathcal{N}_{A'\to B}$, the following is an achievable set of rate-triples (C, Q, E) of classical communication,*

quantum communication, and entanglement consumption:

$$C + 2Q \leq I(AX;B)_\sigma, \tag{34}$$

$$Q + E \leq I(A\rangle BX)_\sigma, \tag{35}$$

$$C + Q + E \leq I(X;B)_\sigma + I(A\rangle BX)_\sigma, \tag{36}$$

where σ is a state of the form $\sigma_{AXB} = \sum_x p_x |x\rangle\langle x|_X \otimes \mathcal{N}_{A'\to B}(\phi^x_{AA'})$ for pure states $\phi^x_{AA'}$.

If we denote this region as $\mathcal{C}^{(1)}_{CQE,\sigma}(\mathcal{N})$, and set $\mathcal{C}^{(1)}_{CQE}(\mathcal{N}) := \bigcup_\sigma \mathcal{C}^{(1)}_{CQE,\sigma}(\mathcal{N})$, then the dynamic capacity region is the closure of its regularization: $\mathcal{C}_{CQE}(\mathcal{N}) = \overline{\bigcup_{k=1}^\infty \frac{1}{k}\mathcal{C}^{(1)}_{CQE}(\mathcal{N}^{\otimes k})}$.

This is the generalization of the unit resource capacity region (Theorem 6) when the sender and receiver have access to many independent uses of a noisy quantum channel $\mathcal{N}$. As before, given a triple (C, Q, E), a positive value for C, Q, or E indicates that the corresponding resource is produced, whereas a negative value means the resource is consumed.

VIII. Capacity of Optical Communications

Our next focus will be to examine the dynamic capacity region in the regime of optical communications.

A. Quantum Optical States

We begin with a necessary review of the basics of optical quantum states, which arise from quantizing the electromagnetic field. They live in bosonic fields which can be quantized similarly to the quantum harmonic oscillator. Optical states travel in spatial modes which are either transverse or longitudinal to their direction of propagation. Orthogonal transverse modes can be perfectly distinguished by a receiver, so each transverse mode corresponds to a single communication channel.

For each mode in a bosonic field, we have an associated operator, $\hat{a}$, referred to as the "destruction" or "annihilation" operator. Its conjugate $\hat{a}^\dagger$ is called the "creation" operator. We can decompose $\hat{a}$ into its real and imaginary parts: $\hat{a} = \hat{a}_1 + j\hat{a}_2$, where $\hat{a}_1 = \frac{1}{\sqrt{2}}(\hat{a} + \hat{a}^\dagger)$ and $\hat{a}_2 = \frac{j}{\sqrt{2}}(\hat{a} - \hat{a}^\dagger)$. We call $\hat{a}_1$ and $\hat{a}_2$ the real and imaginary *quadrature* operators, which are Hermitian observables. The annihilation and creation operators satisfy the fundamental commutation relation $[\hat{a}, \hat{a}^\dagger] = 1$, and the quadratures satisfy $[\hat{a}_1, \hat{a}_2] = \frac{j}{2}$. From the Heisenberg Uncertainty Principle, this implies that for any quantum state $|\psi\rangle$, we have $\langle\Delta\hat{a}_1^2\rangle\langle\Delta\hat{a}_2^2\rangle \geq 1/16$. Here, following convention, we define for an operator $\hat{A}$ and a state $|\psi\rangle$ the expected value $\langle\hat{A}\rangle = \langle\psi|\hat{A}|\psi\rangle$ and variance $\langle\Delta\hat{A}\rangle = \langle\hat{A} - \langle\hat{A}\rangle\rangle$. States whose quadrature components achieve the lower bound in the

Heisenberg Uncertainty Principle are called *quadrature minimum-uncertainty* states, or *minimum uncertainty product* (MUP) states.

In a closed system, the annihilation operator evolves as a function of time, $\hat{a}(t)$, which is governed by the Heisenberg Equation of Motion: $j\hbar\frac{d\hat{a}(t)}{dt} = [\hat{a}(t), \hat{H}] = \hbar\omega\hat{a}(t)$, where ω is the frequency of the mode and $\hat{H}$ is the Hamiltonian of the system, an observable representing the system's energy. It follows that $\hat{a}(t) = \hat{a}e^{-j\omega t}$ and $\hat{H} = \hbar\omega[\hat{a}^\dagger\hat{a}+1/2] = \hbar\omega[\hat{a}_1^2(t) + \hat{a}_2^2(t)]$.

1. Number States

The product $\hat{N} := \hat{a}^\dagger\hat{a}$ appearing in the Hamiltonian is called the "number" operator, so named because it has an orthogonal set of eigenvectors $\{|n\rangle\}$ indexed by the nonnegative integers $n \geq 0$, where $\hat{N}|n\rangle = n|n\rangle$. These discrete eigenvalues correspond to discrete energy levels of the Hamiltonian, which in turn correspond to the number of photons in the state. We call these *Fock states*, or simply *number* states. The number states form a complete basis ($I = \sum_{n=0}^{\infty} |n\rangle\langle n|$), and each state $|n\rangle$ contains exactly n photons. We call the state $|0\rangle$ the *vacuum state*, as it contains no photons. The annihilation and creation operators act respectively on number states as

$$\hat{a}|n\rangle = \sqrt{n}|n-1\rangle \text{ (and } \hat{a}|0\rangle = |0\rangle), \tag{37}$$

$$\hat{a}^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle, \tag{38}$$

which explains where these operators get their names. With respect to a number state $|n\rangle$, the quadrature components have mean $\langle\hat{a}_1\rangle = \langle\hat{a}_2\rangle = 0$ and variance $\langle\Delta\hat{a}_1\rangle = \langle\Delta\hat{a}_2\rangle = \frac{2n+1}{4}$, so $|n\rangle$ is only a quadrature minimum-uncertainty state when $n = 0$.

2. Coherent States

Another important state is called a *coherent* state $|\alpha\rangle$, which is an eigenvector of the annihilation operator: $\hat{a}|\alpha\rangle = \alpha|\alpha\rangle$. There is a coherent state for every complex number $\alpha \in \mathbb{C}$, and expressed in the basis of number states we have

$$|\alpha\rangle = \sum_{n=0}^{\infty} \frac{\alpha^n e^{-|\alpha|^2/2}}{\sqrt{n!}}|n\rangle. \tag{39}$$

Coherent states are not orthogonal, so they cannot be unambiguously distinguished with a quantum measurement. But they are a complete set since they resolve the identity operator: $I = \int \frac{d^2\alpha}{\pi}|\alpha\rangle\langle\alpha|$. As a result, we may write $\hat{a} = \int \frac{d^2\alpha}{\pi}\alpha|\alpha\rangle\langle\alpha|$. The probability of observing n photons in the state $|\alpha\rangle$ by measuring $\hat{N}$ is $\frac{|\alpha|^{2n}}{n!}e^{-|\alpha|^2}$, which is a Poisson distribution with mean and variance equal to $|\alpha|$. With respect to $|\alpha\rangle$ where $\alpha = \alpha_1 + j\alpha_2$, the quadrature components have means $\langle\hat{a}_1\rangle = \alpha_1$ and $\langle\hat{a}_2\rangle = \alpha_2$, and variances $\langle\Delta\hat{a}_1\rangle = \langle\Delta\hat{a}_2\rangle = \frac{1}{4}$. Thus, coherent states are quadrature minimum-uncertainty states.

The vacuum state is actually both a number state and a coherent state, so there is no ambiguity in writing it as $|0\rangle$. Any coherent state can be realized by applying a

displacement operator

$$D(\hat{a}, \alpha) = \exp(\alpha \hat{a}^\dagger - \alpha^* \hat{a}) \tag{40}$$

to the vacuum state:

$$|\alpha\rangle = D(\hat{a}, \alpha)|0\rangle. \tag{41}$$

3. Squeezed States

One other important class of states is that of *squeezed* states. If we define the squeeze operator

$$\hat{S}(r, \phi) = \exp\left(\frac{1}{2} r(\hat{a}^2 e^{-2j\phi} - \hat{a}^{\dagger 2} e^{2j\phi})\right), \tag{42}$$

then a quadrature-squeezed state is a state of the form

$$|\alpha\rangle_{(r,\phi)} = D(\hat{a}, \alpha)\hat{S}(r, \phi)|0\rangle. \tag{43}$$

We may set $\phi = 0$ without loss of generality by choosing the phase of our optical frequency appropriately. With respect to a squeezed state $|\alpha\rangle_{(r,\phi)}$, where $\alpha = \alpha_1 + j\alpha_2$, the means of the quadratures are $\langle \hat{a}_1 \rangle = \alpha_1$ and $\langle \hat{a}_2 \rangle = \alpha_2$, and their variances are $\langle \Delta \hat{a}_1 \rangle = \frac{1}{4} e^{-2r}$ and $\langle \Delta \hat{a}_2 \rangle = \frac{1}{4} e^{2r}$. These form all of the quadrature minimum-uncertainty states, with coherent states arising as a subset of the squeezed states when $r = 0$. The term "squeezed" indicates that we have reduced the variance in one of the quadratures at the cost of increasing the variance in the other. This could allow us to modulate information in the squeezed quadrature.

B. Free-space Optical Communication: The Lossy Bosonic Channel

Free-space optical communication is modeled by the *lossy bosonic channel* which, in its most general form, acts on input quantum states as well as the environment. To describe this channel for a single mode, we let $\hat{a}$ be the input annihilation operator for the sender, and $\hat{e}$ the input annihilation operator of the environment. The lossy bosonic channel transforms these operators as

$$\begin{aligned} \hat{a} &\to \sqrt{\eta}\ \hat{a} + \sqrt{1-\eta}\ \hat{e} \\ \hat{e} &\to -\sqrt{1-\eta}\ \hat{a} + \sqrt{\eta}\ \hat{e}. \end{aligned} \tag{44}$$

The constant $\eta \in [0, 1]$ is called the *transmissivity*, and loosely corresponds to the fraction of the sender's input photons which reach the receiver. We obtain a Kraus map $\mathcal{N}$ for this channel by tracing out the environment. It acts on coherent states by attenuating them: $|\alpha\rangle \mapsto \sqrt{\eta}|\alpha\rangle$. When the input environment modes are initially prepared in the vacuum state, we call this a *pure loss channel*, and we refer to it as a *lossless* bosonic channel when $\eta = 1$.

1. Classical Capacity

The classical capacity of the single-mode, single-frequency lossless bosonic channel was studied in [46,47]. For classical information transfer, we assume that we transmit states ρ_θ with a probability distribution $P(d\theta)$, resulting in the expected density operator $\rho = \int \rho_\theta P(d\theta)$. The states are measured at the receiver with a POVM $\Lambda(x)$, where $\Lambda(x) \succeq 0$, $\forall x$, and $\int \Lambda(x)dx = I$. Let X be the random variable associated to the classical outcome of the POVM measurement. Yuen and Ozawa [46] proved that the mutual information between the (classical) random variables θ and X is bounded as

$$I(\theta; X) \leq H(\rho) - \int H(\rho_\theta)P(d\theta). \tag{45}$$

The proof relied on the Uhlmann inequality $D(\mathcal{N}(\sigma_1)||\mathcal{N}(\sigma_2)) \leq D(\sigma_1||\sigma_2)$, applied to the map $\mathcal{N}(\sigma) = \mathrm{Tr}\,(\Lambda(dx)\sigma)$. For a fixed energy constraint $\int \mathrm{Tr}\left(\hat{a}^\dagger \hat{a}\rho_\theta\right) P(d\theta) \leq N$, the right-hand side of (45) is maximized by choosing the alphabet of θ to be the natural numbers and the states $\rho_n = |n\rangle\langle n|$ with the probability distribution $P(n) = N^n(1+N)^{-(n+1)}$. Note that $H(\rho_n) = 0$ for each n. Then, choosing the POVM $\{\Lambda(n)\}$ to be that given by measuring the number operator $\hat{N} = \hat{a}^\dagger\hat{a} = \sum_{n=0}^{\infty} n|n\rangle\langle n|$, the mutual information $I(\theta; X)$ can be shown to be equal to this upper bound, which is therefore the capacity. A direct calculation gives that the capacity of the lossless bosonic channel with energy limit N is given by $C(N) = g(N)$, where $g(x)$ is the function

$$g(x) := (x+1)\log_2(x+1) - x\log_2 x. \tag{46}$$

$g(N)$ is the von Neumann entropy of the expected density operator

$$\rho = \sum_{n=0}^{\infty} \frac{N^n}{(N+1)^{n+1}}|n\rangle\langle n|, \tag{47}$$

which is called a *thermal state* with mean photon number N.

Giovannetti et al. [48] were able to generalize this result to arbitrary lossy bosonic channels. We denote a single-mode channel as $\mathcal{N}$, and if we allow ourselves n successive channel uses, we effectively form the channel $\mathcal{N}^{\otimes n}$. In the case of multimode channels, we may symbolize the channel corresponding to the k^{th} mode as $\mathcal{N}_k$ (which may be comprised of successive channel uses as above), and denote the full channel as $\bigotimes_k \mathcal{N}_k$. If ω_k and N_k are the frequency and average photon number of the k^{th} mode, then we assume an energy constraint of the form $\sum_k \hbar\omega_k N_k = E$.

From the HSW theorem, the capacity for the single-mode channel is $C = \sup_n \frac{C_n}{n}$ where, like before,

$$C_n := \max_{\{P(\theta),\rho_\theta\}} H(\sigma) - \int H(\sigma_\theta)P(\theta)d\theta, \tag{48}$$

where $\rho_\theta \in \mathcal{H}^{\otimes n}$, $\sigma_\theta = \mathcal{N}^{\otimes n}(\rho_\theta)$, $P(\theta)$ is a probability density function, and $\sigma = \int \sigma_\theta P(\theta)d\theta$. [48] showed that it is possible to achieve C as the Holevo information of an ensemble of states with just $n = 1$ parallel channel uses. A capacity-achieving ensemble

is shown to be Gaussian mixture of coherent states in each mode. In mode k, we select the alphabet of θ to be the complex numbers, and choose an ensemble of the form $\{P_k(\alpha), |\alpha\rangle_k\}$, $\alpha \in \mathbb{C}$, where

$$P_k(\alpha) = \frac{1}{\pi N_k} \exp[-|\alpha|^2/N_k]. \tag{49}$$

Under these ensembles, the input state to the entire channel becomes $\bigotimes_k \int |\alpha\rangle\langle\alpha|_k P_k(\alpha) d^2\alpha$. The capacity achieved using this scheme is $g(\eta_k N_k)$ for each mode, where $g(x)$ is the function from (46), and for the entire channel $\bigotimes_k \mathcal{N}_k$ the capacity becomes $\sum_k g(\eta_k N_k)$. Thus, the maximum capacity achievable is

$$C = \max_{\{N_k\}} \sum_k g(\eta_k N_k), \tag{50}$$

where the maximization is with respect to photon numbers N_k which satisfy the energy constraint. [48] actually shows that Equation (50) is an upper bound on $\frac{C_n}{n}$ for all n, which implies that the capacity can be achieved by taking $n = 1$, i.e. the capacity is single-letter. The above-mentioned capacity-achieving input state does not require entanglement between different modes. Note that the proof of the HSW theorem involves forming codewords from sequences of states drawn from the ensemble $\{P_k(\alpha), |\alpha\rangle_k\}$, so it is possible that exploiting entanglement at the sender's end can have benefits such as reducing the error probability for finite-length block codes.

2. Entanglement-Assisted Classical Capacity

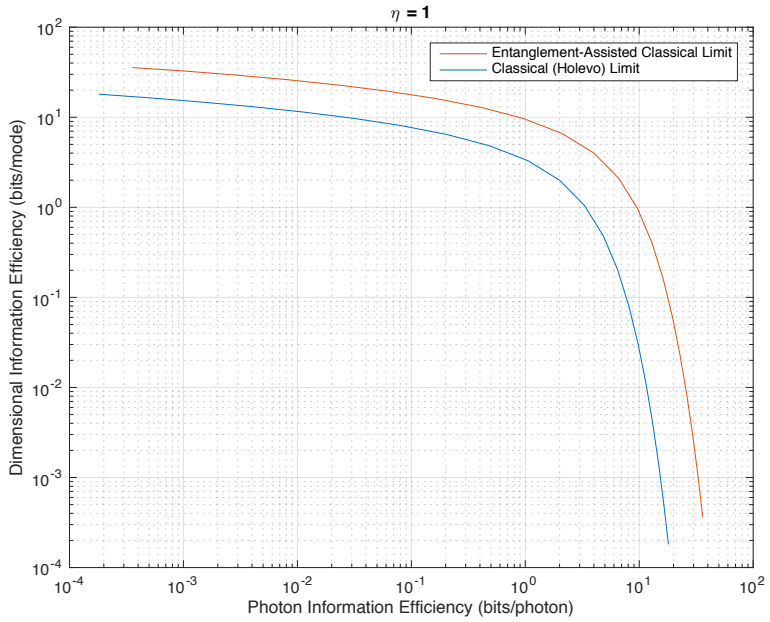
Giovannetti et al. characterized the entanglement-assisted classical capacity of the lossy bosonic channel in [49]. For a single-mode lossy bosonic channel $\mathcal{N}$ with transmissivity η and average photon number N, this capacity turns out to be

$$C_E(\mathcal{N}) = g(\eta N) + g(N) - g((1-\eta)N) \text{ bits per channel use.} \tag{51}$$

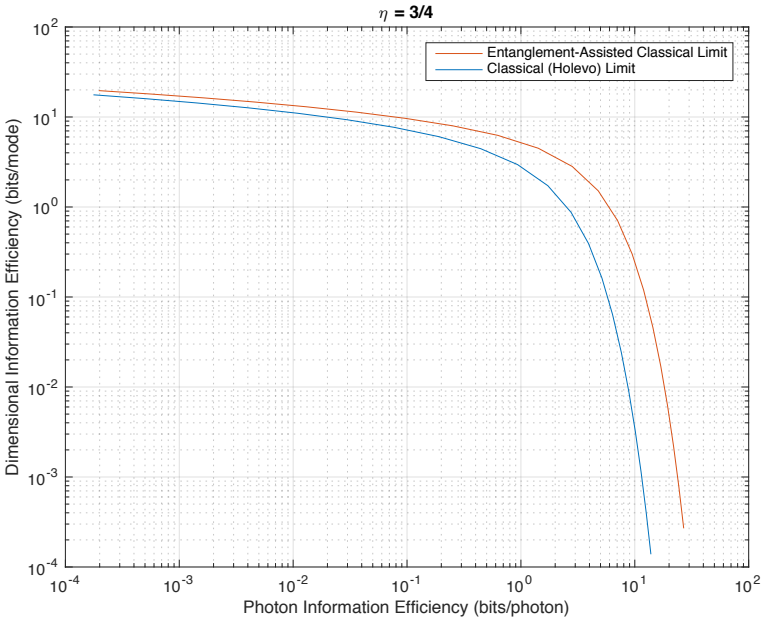
Contrast this to the classical capacity of $C(\mathcal{N}) = g(\eta N)$ bits per channel use. The entanglement-assisted classical capacity can be achieved by selecting coherent states randomly from the same ensemble $\{P(\alpha), |\alpha\rangle\}$ as in the classical-capacity-achieving case, where $P(\alpha)$ is defined as in (49). For a multimode channel $\bigotimes_k \mathcal{N}_k$, with transmissivities $\{\eta_k\}$, frequencies ω_k, and average photon numbers N_k which must satisfy $\sum_k \hbar\omega_k N_k = E$, the entanglement-assisted classical capacity of the resulting channel will be

$$C_E\left(\otimes_k \mathcal{N}_k\right) = \max_{\{N_k\}} \sum_k g(\eta_k N_k) + g(N_k) - g((1-\eta_k)N_k). \tag{52}$$

Figure 1 illustrates the capacity gain from using entanglement-assistance over the single-mode lossy bosonic channel. The plots show trade-off curves for the maximum number of classical bits which can be sent per photon versus per mode for transmissivities $\eta = 1$ and $\eta = 3/4$. As the we can see from the plots, the maximum improvement we can attain in dimensional information efficiency while maintaining a constant photon information efficiency by using entanglement assistance depends largely on the value of η. The same is true if we consider the improvement in photon information efficiency at a fixed dimensional information efficiency.



(a)



(b)

Figure 1. The improvement in classical capacity using entanglement-assistance for the single-mode bosonic channel with transmissivity η, where $\eta = 1$ in 1(a) and $\eta = \frac{3}{4}$ in 1(b) . Dimensional information efficiency is equivalent to the capacity per channel use, and photon information efficiency is this capacity divided by the average number of photons per mode, N.

3. Quantum Capacity

The foundations for proving the quantum capacity of the lossy bosonic channel were set in [50], and solidified in [51, 52]. Perhaps unsurprisingly, for a single-mode channel $\mathcal{N}$ with transmissivity η and average photon number N, the quantum capacity has a form similar to those of the classical and entanglement-assisted classical capacities. The quantum capacity is

$$Q(\mathcal{N}) = g(\eta N) - g((1-\eta)N). \tag{53}$$

The results of [49] and [53] further develop this result by finding the entanglement-assisted quantum capacity:

$$Q_E(\mathcal{N}) = \frac{1}{2}[g(\eta N) + g(N) - g((1-\eta)N)]. \tag{54}$$

Both the quantum and entanglement-assisted quantum capacities can be achieved by generating random quantum codes from a thermal state distribution as in the expected density operator of Equation (47).

4. The Dynamic Capacity Region

To conclude, we discuss the dynamic capacity region of the lossy bosonic channel—Equations (34), (35), and (36) of Theorem 16. This region was characterized in [53], which found that for a channel with transmissivity η and a mean photon number of N_S, these three equations take the form:

$$C + 2Q \leq g(\lambda N_S) + g(\eta N_S) - g((1-\eta)\lambda N_S), \tag{55}$$

$$Q + E \leq g(\eta\lambda N_S) - g((1-\eta)\lambda N_S), \tag{56}$$

$$C + Q + E \leq g(\eta N_S) - g((1-\eta)\lambda N_S), \tag{57}$$

which is a family of achievable regions of rate-triples (C, Q, E) parametrized by $\lambda \in [0, 1]$. Each such region comes from employing a coding ensemble of states constructed from displacements of the *two-mode squeezed vacuum state*:

$$|\psi_{TMS}\rangle_{AA'} := \sum_{n=0}^{\infty} \sqrt{\frac{(\lambda N_S)^n}{(\lambda N_S + 1)^{n+1}}} |n\rangle_A |n\rangle_{A'}. \tag{58}$$

The ensemble consists of sending the state $D(\hat{a}', \alpha)|\psi_{TMS}\rangle$ with probability

$$p_\alpha = \frac{1}{\pi(1-\lambda)N_S} \exp\left(\frac{-|\alpha|^2}{(1-\lambda)N_S}\right), \tag{59}$$

where $\hat{a}'$ is the annihilation operator of the A' system and $D(\hat{a}', \alpha)$ is the displacement operator from Equation (40). In the language of Theorem 16, the states are now parametrized by the complex number α, with "p_x" becoming "p_α" and "$\phi^x_{AA'}$" becoming "$D(\hat{a}', \alpha)|\psi_{TMS}\rangle_{AA'}$."

When we take the parameter $\lambda = 0$, the ensemble becomes $\{p(\alpha), |0\rangle_A \otimes |\alpha\rangle_{A'}\}$, the Gaussian mixture of coherent states which achieves the classical capacity of the lossy bosonic channel as per [48]. When $\lambda = 1$, we recover the thermal state distribution which achieves the quantum capacity of the channel. This reflects the nature of the protocol constructed in [53] to achieve the dynamic capacity region. Loosely speaking, λ represents the fraction of photons dedicated to the "quantum" part of the code.

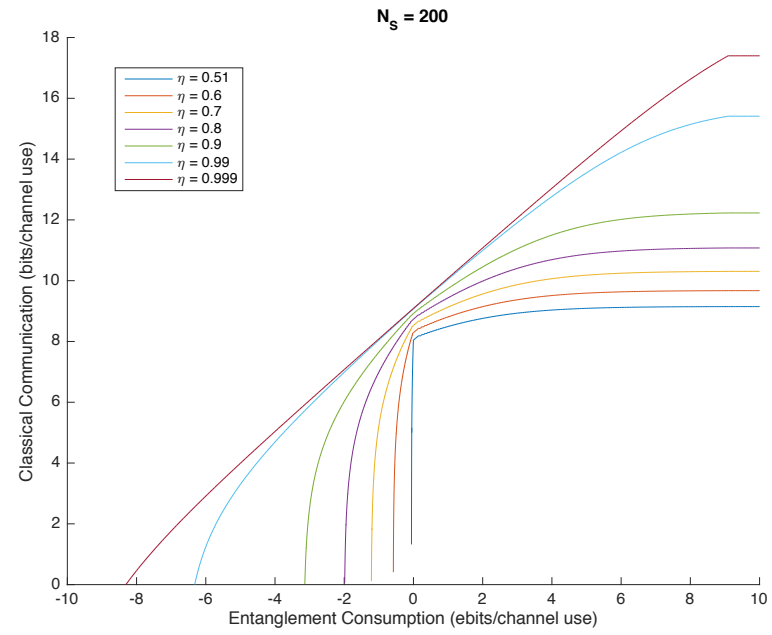
Figure 2 illustrates the trade-off between the rates of entanglement consumption E and classical communication C implied by Equations (55), (56), and (57). The fact that C is a concave function of E shows that we can outperform timesharing between two protocols for consuming ebits to communicate classical bits. It also suggests that given a fixed entanglement budget of N_e ebits for a given number n of channel uses, a strategy for maximizing the number of bits sent would amount to using the same number of ebits in each channel use—that is, consume N_e/n ebits per channel use. If, however, the demand for classical information is lower at a particular channel use, we can consume a smaller number of ebits to send the required number of bits. In fact, if we can afford a high enough mean photon number N_S, we can actually *generate* new ebits of entanglement between the sender and receiver, which we can use in a later channel use should we need to communicate at a higher classical capacity. If we consider the cumulative number of ebits at our disposal as an "entanglement battery," this setup reflects a power network problem.

C. Approaching Capacity in Practice

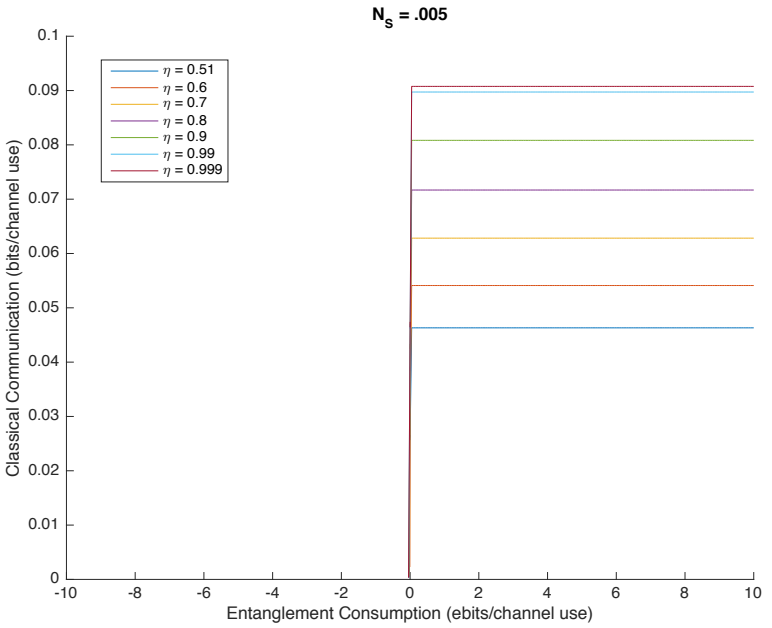
We now discuss how high a capacity we can achieve on the bosonic channel using conventional methods. We will restrict our attention to the case of perfect transmissivity: $\eta = 1$. Practical methods for communication over a single transverse mode typically involve selecting a corresponding set of orthogonal longitudinal modes as well as a set of quantum states to transmit over each mode. The overall quantum state over a time interval will be the product state of the quantum states for each mode. Examples include pulse position modulation (PPM), in which the orthogonal modes are disjoint pulses in time, and frequency multiplexing, in which the modes correspond to dividing the available bandwidth into disjoint frequency bins.

1. The Number State Channel

Let us first examine communicating over one such bin using Fock states—the narrowband single number-state channel. Here, information is modulated in the form of number states $|n\rangle$, and an ideal photodetector is used at the receiver to count the number of photons in each orthogonal mode. We communicate at an operating frequency f and a bandwidth $B << f$, and we model the channel as having thermal noise: For any channel use, the photodetector could falsely detect an extra k photons with probability $q(k) = \frac{1}{1+\bar{n}_T}\left(\frac{\bar{n}_T}{1+\bar{n}_T}\right)^k$, where $\bar{n}_T = \frac{1}{e^{hf/k_B T}-1}$ with k_B the Boltzman constant and T the temperature. If we are allowed an average photon number $\bar{n} := \sum_n p_n n$ for our



(a)



(b)

Figure 2. The optimal trade-off between the rate of entanglement consumption (E) and classical communication (C) over the lossy bosonic channel with transmissivity η ranging between $.5$ and 1. Negative values of entanglement consumption indicate generated entanglement between sender and receiver. Plots are shown for mean photon numbers of $N_S = 200$ in (a) and $N_S = .005$ in (b).

number states, as determined by a power constraint $P = Bhfn$, then [47] shows that the capacity-achieving input distribution is $\{p_n, |n\rangle\}$ with

$$p_n = \begin{cases} \frac{1+\bar{n}_T}{1+\bar{n}+\bar{n}_T}, & n = 0 \\ \frac{\bar{n}}{(\bar{n}+\bar{n}_T)(1+\bar{n}+\bar{n}_T)} \left(\frac{\bar{n}+\bar{n}_T}{1+\bar{n}+\bar{n}_T}\right)^n, & n > 0 \end{cases}. \tag{60}$$

The capacity under this distribution becomes

$$\mathcal{C}_T = (\bar{n} + \bar{n}_T) \log_2 \left(1 + \frac{1}{\bar{n} + \bar{n}_T}\right) + \log_2 \left(1 + \frac{\bar{n}}{1 + \bar{n}_T}\right) - \bar{n}_T \log_2 \left(1 + \frac{1}{\bar{n}_T}\right). \tag{61}$$

This is maximal when $T = 0$, in which case p_n becomes a thermal distribution and the capacity is simply

$$\mathcal{C}_0 = \bar{n} \log_2(1 + \bar{n}^{-1}) + \log_2(1 + \bar{n}). \tag{62}$$

The capacity per unit time is

$$C_0 = B\mathcal{C}_0 = B[\bar{n} \log_2(1 + \bar{n}^{-1}) + \log_2(1 + \bar{n})]. \tag{63}$$

[47] then considers the regime in which we fix the fractional bandwidth $\gamma = B/f << 1$, and attempt to maximize C_0 by varying f (which is equivalent to varying $\bar{n}$ given a fixed power $P = Bhf\bar{n}$). The maximum capacity turns out to be $C_{max} = 2\sqrt{\gamma h/P}$ bits per second, which is achieved when $\bar{n} = 1$ (or equivalently, $f = \sqrt{P/\gamma h}$). This corresponds to $C_{max}/B\bar{n} = 2$ bits per photon. Finally, the results are generalized to the wideband case by using frequency-multiplexing, in which the total bandwidth is divided into bins of size b, with bin i centered at frequency f_i and with average photon number $\bar{n}_i$. Our power constraint now takes the form

$$P = b \sum_i h f_i \bar{n}_i, \tag{64}$$

and we maximize the overall information rate with respect to $\bar{n}_i$ (as a function of f_i) subject to constant power P. The resulting wideband capacity is shown to be $C_{WB} = \frac{\pi}{\ln 2}\sqrt{\frac{2P}{3h}}$ bits/s, with the average photon numbers taking the form $\bar{n}_i = \frac{1}{e^{\beta h f_i} - 1}$ where β is determined from power constraint.

2. The Coherent State Channel

We next consider using coherent states for communication, modulating information in both quadrature components and using *heterodyne detection* to measure both components together. More formally, the coherent state $\hat{\rho}_\alpha = |\alpha\rangle\langle\alpha|$ is transmitted according to a probability density function $p(\alpha)$, leading to an expected density operator $\hat{\rho} = \int p(\alpha)\hat{\rho}_\alpha d^2\alpha$ with respect to the measure $d^2\alpha = d\alpha_1 d\alpha_2$ where $\alpha = \alpha_1 + j\alpha_2$. Ideal heterodyne detection measures the POVM $\{\Lambda_\beta\}$ where $\Lambda_\beta := \frac{1}{\pi}|\beta\rangle\langle\beta|$.

In the narrowband scenario, we have a single mode of frequency ω. A power constraint becomes a constraint on the mean number of photons per channel use, $\bar{n} = \mathrm{Tr}\left(\hat{\rho}\hat{a}^\dagger\hat{a}\right) =$

$\int p(\alpha)|\alpha|^2 d^2\alpha$. It is shown in [47] that the detection error is equivalent to additive Gaussian noise, from which it is deduced that the capacity-achieving distribution is given by the Gaussian input density $p(\alpha) = \frac{1}{\pi\bar{n}} \exp\left(-\frac{|\alpha|^2}{\bar{n}}\right)$, which makes the expected density operator $\hat{\rho}$ a thermal state. The resulting capacity per channel use is calculated to be $\mathcal{C} = \log_2(1+\bar{n})$. Hall (1993) generalized this result to the coherent state channel with further additive Gaussian noise.

As in the number-state channel, we can again fix the fractional bandwidth $\gamma = B/f$ and vary $f\bar{n}$ to maximize the capacity per unit time, C. This capacity is found to be $C_{max} = 1.1610\sqrt{\frac{\gamma P}{h}}$ bits per second, or 0.58628 bits per photon, achieved when $\bar{n} = 3.9216$ (see [47]).

We can also address the wideband case, as before, assuming a zero-temperature frequency-multiplexed channel where we again divide the bandwidth into equal bins of size b, where $\bar{n}_i$ is the mean photon number of the i^{th} bin. Our power constraint again takes the form from Equation (64), and the capacity now becomes $C = b\sum_i \log_2(1+\bar{n}_i)$. We maximize this with respect to the $\bar{n}_i$, each a function of the corresponding bin frequency f_i through the fixed power constraint. Communication becomes too inefficient above $f_c = \sqrt{\frac{2P}{h}}$ [47], leading to an optimal mean photon number of

$$\bar{n}_i = \begin{cases} \frac{f_c}{f_i} - 1, & 0 \le f \le f_c, \\ 0, & f \ge f_c \end{cases} \tag{65}$$

and a wideband capacity of $C_{WB} = \frac{f_c}{\ln 2} = \frac{1}{\ln 2}\sqrt{\frac{2P}{h}}$ bits/s.

3. The Quadrature-Squeezed Channel

Suppose now that we modulate information in a single quadrature of a squeezed state, and the receiver attempts to measure the single quadrature using *homodyne detection*. In particular, the sender transmits states of the form $\hat{\rho}_{\alpha_1} = |\alpha_1\rangle_{(r,0)(r,0)}\langle\alpha_1|$, where α_1 is a real number corresponding to a measurement of the real quadrature component $\hat{a}_1$ of $\hat{a} = \hat{a}_1 + j\hat{a}_2$. Now information is modulated only on $\hat{a}_1$, whose noise is squeezed below the level of a typical coherent state. We analyze the case where the squeeze parameter r is the same for all sent states, though we could potentially do better with r as a function of a_1. We also choose $\langle\hat{a}_2\rangle = \alpha_2 = 0$ so as not to waste energy, though squeezing $\hat{a}_1$ quadrature will increase noise of $\hat{a}_2$ and consume power. Our expected density operator $\hat{\rho} = \int p(\alpha_1)\hat{\rho}_{\alpha_1} d\alpha_1$ is now with respect a probability density operator $p(\alpha_1)$ on the real numbers. Our mean photon number becomes $\bar{n} = \mathrm{Tr}\left(\hat{\rho}\hat{a}^\dagger\hat{a}\right) = \sigma^2 + \sinh^2 r$, where $\sigma^2 = \int p(\alpha_1)\alpha_1^2 d\alpha_1$, which we assume is fixed as we are operating at a single mode of frequency ω with a power constraint.

Ideal homodyne detection corresponds to measuring the POVM $\{\Lambda_{x_1}\}$, where $\Lambda_{x_1} = |x_1\rangle\langle x_1|$. Once again, [47] argues that the detection error takes the form of additive Gaussian noise, with the resulting capacity-achieving input distribution given by the

Gaussian density function $p(\alpha_1) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\alpha_1^2}{2\sigma^2}\right)$. The maximal capacity per channel use is then $\mathcal{C} = \log_2(1 + 2\bar{n})$, which is achieved by choosing r to satisfy $e^{2r} = \log_2(1 + 2\bar{n})$.

In this case, when the fractional bandwidth is fixed, the maximum capacity per unit time is found to be $C_{max} = 1.1610\sqrt{\frac{2\gamma P}{h}}$ bits per second (1.17256 bits per photon) achieved when $\bar{n} = 1.9608$ (see [47]).

In the wideband regime, again using the same frequency-multiplexing argument from before, the capacity per unit time takes the form $C = b\sum_i \log_2(1+2\bar{n}_i)$. Communication is now too inefficient above the frequency $\tilde{f}_c = \sqrt{\frac{4P}{h}}$. The optimal mean photon number for bin i is

$$\bar{n}_i = \begin{cases} \frac{\tilde{f}_c}{2f_i} - 1, & 0 \leq f \leq \tilde{f}_c, \\ 0, & f \geq \tilde{f}_c, \end{cases} \tag{66}$$

and the wideband capacity is $C_{WB} = \frac{\tilde{f}_c}{\ln 2} = \frac{1}{\ln 2}\sqrt{\frac{4P}{h}}$ bits per second [47].

IX. Summary and Future Work

In this article, we reviewed the capacity of a quantum channel for transmitting classical and quantum information. We discussed the necessary trade-offs between the rates of sending both bits and qubits over a quantum channel, and how these rates are affected in the presence of shared entanglement between sender and receiver. We also reviewed several protocols which achieve these rates. Then, narrowing our focus to the free-space optical channel, we reviewed common quantum states of photons and methods for modulating information on them. We discussed the capacities associated with these modulation techniques as well as the prospect of feasibly generating and exploiting entanglement to boost the classical capacity of optical free-space communication.

Future theoretical studies over the next several years will explore the evaluation, achievability, and applicability to NASA's communication systems of the four different types of quantum channel capacities. Problems of interest include: designing practical near-optimal ebit-sharing protocols and privacy protocols; quantifying the trade-off entanglement-assisted capacity versus rate of ebit consumption; quantifying the superadditivity of quantum capacity and private classical capacity; exploiting the relationship between quantum computing and quantum communication; designing efficient codes for classical and entanglement assisted optical channels; quantifying how the various capacities change when restricted to a small set of states and POVMs; determining the feasibility of entanglement-manipulation when restricted to lab-implemented quantum transformations; computing the achievable capacity of further protocols such as superdense teleportation [54]; constructing new quantum key-distribution protocols; and designing quantum error-correcting codes to send classical and private information over a quantum channel.

References

[1] M. M. Wilde. *Quantum Information Theory, 2nd edition*. Cambridge University Press, 2016.

[2] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[3] E. B. Davies and J. T. Lewis. An operational approach to quantum probability. *Communications in Mathematical Physics*, 17(3):239–260, 1970.

[4] M. Ozawa. Quantum measuring processes of continuous observables. *Journal of Mathematical Physics*, 25(1):79–87, 1984.

[5] V. Paulsen. *Completely Bounded Maps and Operator Algebras*. Cambridge University Press, 2002.

[6] M. R. Audenaert. A sharp continuity estimate for the von neumann entropy. *Journal of Physics A: Mathematical and Theoretical*, 40(28):8127–8136, 2007.

[7] M. Fannes. A continuity property of the entropy density for spin lattice systems. *Communications in Mathematical Physics*, 31(4):291–294, December 1973.

[8] B. Schumacher. Quantum coding. *Physical Review A*, 51(4):2738–2747, April 1995.

[9] E. H. Lieb and M. B. Ruskai. Proof of the strong subadditivity of quantum-mechanical entropy. *Journal of Mathematical Physics*, 14:1938–1941, 1973.

[10] E. H. Lieb and M. B. Ruskai. A fundamental property of quantum-mechanical entropy. *Physical Review Letters*, 30(10):434–436, March 1973.

[11] A. S. Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Problems of Information Transmission*, 9:177–183, 1973.

[12] M.-D. Choi. Completely positive linear maps on complex matrices. *Linear Algebra and its Applications*, 10:285–290, 1975.

[13] C. H. Bennett and S. J. Wiesner. Communication via one- and two-particle operators on einstein-podolsky-rosen states. *Physical Review Letters*, 69(20):2881–2884, November 1992.

[14] C. H. Bennett, G. Brassard, C. Crepeau, R. Jozsa, A. Peres, and W. K. Wootters. Teleporting an unknown quantum state via dual classical and einstein-podolsky-rosen channels. *Physical Review Letters*, 70(13):1895–1899, March 1993.

[15] M.-H. Hsieh and M. M. Wilde. Trading classical communication, quantum communication, and entanglement in quantum shannon theory. *IEEE Transactions on Information Theory*, 56(9):4705–4730, September 2010.

[16] A. S. Holevo. The capacity of the quantum channel with several general signal states. *IEEE Transactions on Information Theory*, 44(1):269–273, January 1998.

[17] B. Schumacher and M. D. Westmoreland. Sending classical information via noisy quantum channels. *Physical Review A*, 56(1):131–138, July 1997.

[18] M.-H. Hsieh, I. Devetak, and A. Winter. Entanglement-assisted capacity of quantum multiple-access channels. *IEEE Transactions on Information Theory*, 54(7):3078–3090, July 2008.

[19] M. B. Hastings. Superadditivity of communication capacity using entangled inputs. *Nature Physics*, 5:255–257, April 2009.

[20] M. Horodecki, P. W. Shor, and M. B. Ruskai. Entanglement breaking channels. *Reviews in Mathematical Physics*, 15(6):629–641, 2003.

[21] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal. Entanglement-assisted classical capacity of noisy quantum channels. *Physical Review Letters*, 83(15):3081–3084, October 1999.

[22] I. Devetak. The private classical capacity and quantum capacity of a quantum channel. *IEEE Transactions on Information Theory*, 51(1):44–55, January 2005.

[23] N. Cai, A. Winter, and R. W. Yeung. Quantum privacy and quantum wiretap channels. *Problems of Information Transmission*, 40(4):318–336, October 2004.

[24] R. Ahlswede and A. Winter. Strong converse for identification via quantum channels. *IEEE Transactions on Information Theory*, 48(3):569–579, March 2002.

[25] G. Smith, J. M. Renes, and J. A. Smolin. Structured codes improve the bennett-brassard-84 quantum key rate. *Physical Review Letters*, 100(17):170502, April 2008.

[26] K. Li, A. Winter, X. Zou, and G.-C. Guo. Private capacity of quantum channels is not additive. *Physical Review Letters*, 103(12):120501, September 2009.

[27] G. Smith. Private classical capacity with a symmetric side channel and its application to quantum cryptography. *Physical Review A*, 78(2):022306, August 2008.

[28] S. Lloyd. Capacity of the noisy quantum channel. *Physical Review A*, 55(3):1613–1622, March 1997.

[29] P. W. Shor. The quantum channel capacity and coherent information. *MSRI Workshop on Quantum Computation Lecture Notes*, 2002.

[30] D. DiVincenzo, P. W. Shor, and J. A. Smolin. Quantum channel capacity of very noisy channels. *Physical Review A*, 57(2):830–839, 1998.

[31] G. Smith and J. Yard. Quantum communication with zero-capacity channels. *Science*, 321:1812–1816, September 2008.

[32] I. Devetak and P. W. Shor. The capacity of a quantum channel for simultaneous transmission of classical and quantum information. *Communications in Mathematical Physics*, 256(2):287–303, June 2005.

[33] M. Horodecki, P. Horodecki, and R. Horodecki. Separability of mixed states: necessary and sufficient conditions. *Physical Letters A*, 223(1-2):1–8, 1996.

[34] A. Peres. Separability criterion for density matrices. *Physical Review Letters*, 77:1413–1415, 1996.

[35] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher. Concentrating partial entanglement by local operations. *Physical Review A*, 53(4):2046–2052, April 1996.

[36] H-K Lo and S. Popescu. Classical communication cost of entanglement manipulation: Is entanglement an interconvertible resource? *Physical Review Letters*, 83(7):1459–1462, August 1999.

[37] H-L Lo and S. Popescu. Concentrating entanglement by local actions: Beyond mean values. *Physical Review A*, 63(2):022301, January 2001.

[38] P. Hayden and A. Winter. Communication cost of entanglement transformations. *Physical Review A*, 67(1):012326, January 2003.

[39] A. W. Harrow and H-K Lo. A tight lower bound on the classical communication cost of entanglement dilution. *IEEE Transactions on Information Theory*, 50(2):319–327, February 2004.

[40] A. W. Harrow. Coherent communication of classical messages. *Physical Review Letters*, 92(9):097902, March 2004.

[41] I. Devetak, A. W. Harrow, , and A. Winter. A family of quantum protocols. *Physical Review Letters*, 93(23):239503, December 2004.

[42] I. Devetak, A. W. Harrow, and A. Winter. A resource framework for quantum shannon theory. *IEEE Transactions on Information Theory*, 54(10):4587–4618, October 2008.

[43] P. W. Shor. *Quantum Information, Statistics, Probability (Dedicated to A. S. Holevo on the occasion of his 60th Birthday): The classical capacity achievable by a quantum channel assisted by limited entanglement.* Rinton Press, Inc., 2004.

[44] M.-H. Hsieh and M. M. Wilde. Entanglement-assisted communication of classical and quantum information. *IEEE Transactions on Information Theory*, 56(9):4682–4704, September 2010.

[45] M. M. Wilde and M.-H. Hsieh. The quantum dynamic capacity formula of a quantum channel. *Quantum Information Processing*, 11(6):1431–1463, December 2012.

[46] H. P. Yuen and M. Ozawa. Ultimate information carrying limit of quantum systems. *Physical Review Letters*, 70(4):363–366, 1993.

[47] C. M. Caves and P. D. Drummond. Quantum limits on bosonic communication rates. *Reviews of Modern Physics*, 66(2):481–537, 1994.

[48] V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, J. H. Shapiro, and H. P. Yuen. Classical capacity of the lossy bosonic channel: The exact solution. *Physical Review Letters*, 92(2):027902, 2004.

[49] V. Giovannetti, S. Lloyd, L. Maccone, and P. W. Shor. Entanglement assisted capacity of the broadband lossy channel. *Physical Review Letters*, 91(4):047901, 2003.

[50] A. S. Holevo and R. F. Werner. Evaluating capacities of bosonic gaussian channels. *Physical Review A*, 63:032312, 2001.

[51] M. M. Wolf, D. Pérez-García, and G. Giedke. Quantum capacities of bosonic channels. *Physical Review Letters*, 98(13):130501, 2007.

[52] S. Guha, J. H. Shapiro, and B. I. Erkmen. Capacity of the bosonic wiretap channel and the entropy photon-number inequality. *Proceedings of ISIT*, pages 91–95, 2008.

[53] M. M. Wilde, P. Hayden, and S. Guha. Quantum trade-off coding for bosonic communication. *Physical Review A*, 86:062306, 2012.

[54] H. J. Bernstein. SuperDense quantum teleportation. *Quantum Information Processing*, 5(6):451–461, December 2006.